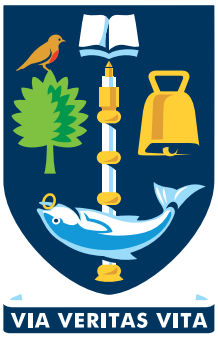
Nanophotonic Filters for Digital Imaging



Kirsty Walls

A thesis submitted to

School of Engineering

University of Glasgow

in fulfilment of the requirements for the degree of

Doctor of Philosophy

June 2013

Preface

This Thesis details the research I conducted at the University of Glasgow as part of the Engineering and Physical Sciences Research Council funded project entitled 'Nanophotonics for Digital Imaging' (EP/G008329/1). The devices detailed in this work were fabricated by Dr Qin Chen, a research associate working on the project, in the James Watt Nanofabrication Centre, University of Glasgow. Aspects of the work described in this Thesis have been published as follows:

International Peer-reviewed Journals

K. Walls, Q. Chen, J. Grant, S. Collins, D. R. S. Cumming and T. D. Drysdale, "Narrowband multispectral filter set for visible band" *Optics Express*, vol. 20, 2012. Also selected to feature in: *Virtual Journal for Biomedical Optics*, vol. 7, no. 11, Oct. 2012

K. Walls, Q. Chen, S. Collins, D. R. S Cumming and T. D. Drysdale, "Automated design, fabrication and characterization of color matching plasmonic filters" *IEEE Photonics Technology Letters*, vol. 24, no. 7, pp. 602-604, April 2012.

Q. Chen, D. Chitnis, **K. Walls**, T. D. Drysdale, S. Collins and D. R. S. Cumming, "CMOS photodetectors integrated with plasmonic color filters," *IEEE Photonics Technology Letters*, vol. 24, no. 3, pp. 197-199, Feb. 2012.

Q. Chen, D. Das, D. Chitnis, **K. Walls**, T. D. Drysdale, S. Collins and D. R. S. Cumming, "A CMOS image sensor integrated with plasmonic colour filters" *Plasmonics* , vol. 7, no. 4, pp. 695-699, Dec. 2012.

International Conference Proceedings

K. Walls, Q. Chen, D. R. S. Cumming and T. D. Drysdale, "Fabry-Pérot resonator with nanostructures for multispectral visible filtering," *12th IEEE Conference on Nanotechnology (IEEE-NANO)*, 20-23 August 2012, Birmingham, UK.

D. R. S. Cumming, Q. Chen, **K. Walls**, T. D. Drysdale, S. Collins, D. Das and D. Chitnis, "Surface plasmon resonance for digital imaging," *12th IEEE Conference on Nanotechnology (IEEE-NANO)*, 20-23 August 2012, Birmingham, UK.

Advanced Training Schools

K. Walls, Q. Chen, D. R. S. Cumming and T. D. Drysdale, "Nanophotonics for digital imaging," (Poster) *Inter-Continental Advanced Materials for Photonics (I-CAMP) Summer School*, 2011, Corrientes, Argentina.

Acknowledgements

First of all, I would like to thank my supervisor, Dr Timothy Drysdale. He has been a constant source of inspiration and sound advice. Thank you for your patience, guidance, motivational talks, and all the effort you have put in to help me become a better engineer.

I would also like to thank my second supervisor Prof David R. S. Cumming for his helpful discussions and guidance throughout my PhD. I am particularly grateful for his efforts in helping me attend ICAMP. I would also like to thank Dr Steve Collins for his helpful advice, and in particular for his suggestions on the development of the CMF filters.

A special acknowledgement goes to Dr Qin Chen, to whom I am particularly indebted for all his work fabricating the devices in this Thesis. I would also like to thank him for his help and guidance at the start of my PhD.

Particular thanks go to Dr James Grant for his help in the fabrication of test structures, and Iain McCrindle for taking the time to review this thesis. Thanks also go to James and Iain for their toleration of my frequent sanity checks and for all their helpful discussions on fabrication, plasmonics and all things research.

I would like to thank all the administrative and technical staff, for their help over the years; for their friendliness, advice and particularly for rescuing me when stranded in foreign lands.

In addition to the above, I would also like to thank all my friends and colleagues who helped make life more fun. Cheers for the coffee, banter and keeping me sane. . . . . just about.

Finally a thank you to my family, Andrea, Shona, Heather and mum for their continuous love, support and patience throughout the years.

Abstract

There has been an increasing demand for low cost, portable CMOS image sensors because of increased integration, and new applications in the automotive, mobile communication and medical industries, amongst others. Colour reproduction remains imperfect in conventional digital image sensors, due to the limitations of the dye-based filters. Further improvement is required if the full potential of digital imaging is to be realised. In alternative systems, where accurate colour reproduction is a priority, existing equipment is too bulky for anything but specialist use.

In this work both these issues are addressed by exploiting nanophotonic techniques to create enhanced trichromatic filters, and multispectral filters, all of which can be fabricated on-chip, i.e. integrated into a conventional digital image sensor, to create compact, low cost, mass produceable imaging systems with accurate colour reproduction.

The trichromatic filters are based on plasmonic structures. They exploit the excitation of surface plasmon resonances in arrays of subwavelength holes in metal films to filter light. The currently-known analytical expressions are inadequate for optimising all relevant parameters of a plasmonic structure. In order to obtain arbitrary filter characteristics, an automated design procedure was developed that integrated a genetic algorithm and 3D finite-difference time-domain tool.

The optimisation procedure's efficacy is demonstrated by designing a set of plasmonic filters that replicate the CIE (1931) colour matching functions, which themselves mimic the human eye's daytime colour response. The best designs were fabricated and demonstrated a least-mean-square error, in comparison to the desired colour matching functions, of 6.37×10^{-3}, 2.34×10^{-3} and 11.10×10^{-3} for the red, green, and blue filters respectively. Notably the spectrum for the red filter contained a double peak, as present in the corresponding colour matching function. Such dual peak behaviour cannot be achieved using a single current dye-based filter. The filters retain the same layer thickness for all structures so

they can be defined in a single lithography step.

A new approach to enable the fabrication of a multispectral filter array on a CMOS imager is also presented. This combines a Fabry-Pérot filter with effective medium theory (EMT) to enable the fabrication of multiple filters in a single cavity length via lithographic tuning of the filter passband. Two approaches are proposed; air-filled nanostructures and dielectric backfilled nanostructures. The air-filled approach is demonstrated experimentally producing three filters with FWHM of $60 - 64\,\text{nm}$. Using the backfilled design, and incorporating a high-index cavity material, a set of twenty three narrowband filters, with a FWHM of $22 - 46\,\text{nm}$ is demonstrated.

A virtual image reproduction process was developed to quantify the image reproduction performance of both the plasmonic and Fabry-Pérot filter sets. A typical rgb dye-based filter set used in conventional imagers achieves a mean colour error of $\Delta E_{00} = 2.711$, whereas the experimental data from the plasmonic filters achieves $\Delta E_{00} = 2.222$ which demonstrated a slight improvement in colour reproduction. The multispectral filter set developed in this work performed even better, with 4 filters giving $\Delta E_{00} = 0.906$, 10 filters $\Delta E_{00} = 0.072$ and continued improvement in the colour error reaching $\Delta E_{00} = 0.047$ for 23 filters. All the filter sets proposed are fully compatible with the CMOS process so as to enable direct integration onto CMOS image sensors in industrial foundries in future. The performance of the presented filters also suggest new compact applications in art reproduction, agricultural monitoring and medical imaging.

Contents

List of Figures

List of Tables

Nomenclature

Greek Symbols

χ	susceptibility
Δ	cell size
δ	phase delay
δ_d	penetration depth into dielectric
δ_m	penetration depth into metal
Γ	reflection amplitude
κ	extinction coefficient of optical material
Λ	period
λ	wavelength
λ_{SPP}	surface plasmon polariton resonant wavelength
μ_r	relative permeability
ρ	reflection coefficient
ε	permittivity

Other Symbols

$\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$	colour matching functions
ΔE_{00}	colour difference derived from the CIEDE2000 system
ΔE_{ab}	colour difference derived from the CIELAB system
$\mathcal{F}$	finesse
$\mathcal{T}$	transmission amplitude
$\tilde{n}$	complex refractive index ($\tilde{n} = n + j\kappa = \sqrt{\varepsilon}$)
$\vec{D}$	electric flux density vector
$\vec{E}$	electric field vector
$\vec{H}$	magnetic field vector

A	transformation matrix
d	diameter
$D(\lambda)$	imaging system response spectrum
F	coefficient of finesse
f	fill factor
k	wavenumber
k_{SPP}	surface plasmon polariton wave vector
l	cavity length
$L(\lambda)$	illuminant spectrum
n	refractive index
N_{sets}	number of filter sets
p_c	probability of crossover
p_m	probability of mutation
R	reflectivity
r, g, b	imager red, green and blue channel signals
$S(\lambda)$	scene spectrum
T	transmittance
t	thickness
D65	CIE standard daylight illuminant
sRGB	standard RGB colour space

Acronyms

$R'G'B'$	gamma corrected RGB
ADC	analogue-to-digital converter
ADE	auxiliary differential equation
ARL	anti-reflection layer
BEOL	back-end-of-line
BSI	back side illumination
CCCR	ColorChecker Color Rendition chart
CCD	charge coupled device
CFA	colour filter array

CIE Commission Internationale de l'Éclairage (International Commission on Illumination)

CIELAB CIE Lab colour space

CIEXYZ tristimulus values derived using the CMF

CMF colour matching functions

CMOS complementary-metal oxide-semiconductor

D2CP Drude + 2 critical points

EB electron beam

EMT effective medium theory

FDTD finite-difference time-domain

FIB focused-ion-beam

FP Fabry-Pérot

FSR free spectral range

FWHM full-width half-maximum

GA genetic algorithm

IR infra-red

ITRS International Technological Roadmap for Semiconductors

LMS least mean square

PMMA poly(methyl methacrylate)

RC recursive convolution

SEM scanning electron micrograph

SPP surface plasmon polariton

TE transverse electric

TEMPEST FDTD code from University of California, Berkeley

TM transverse magnetic

XYZ tristimulus values derived from the rgb device dependent signals

Chapter 1

Introduction

1.1 Motivation

Colour imaging systems are ubiquitous, and can be found in applications as diverse as mobile phone cameras and space exploration. The continued advance of complementary metal-oxide semiconductor (CMOS) fabrication processes, utilised in the fabrication of CMOS image sensors, has resulted in the reduction in the size and cost of imaging technology. With this advancement, imaging systems, or imagers as they are otherwise known, have become more accessible and thus the number of applications has expanded. Two industries where the uptake and integration of imagers has been particularly prevalent are the mobile communication and automotive industries [1]. This growth in the imaging market has resulted in the sales of CMOS image sensors reaching a new record high in 2011,[1] with $5.1 bn USD in sales, accounting for 60% of the market. This is forecast to increase to $7.6 bn and 66% in 2015 [2]. With this increase in the incorporation of imaging systems within such a variety of industries and applications, there is an even greater drive for continued development and miniaturisation. Image sensor manufacturers aim to produce more compact devices with improved resolution and colour reproduction, in addition to increased on-chip capacity. This must be achieved while simultaneously maintaining or improving cost and imaging performance.

[1] the most recent year for which figures are available

Despite the widespread use of CMOS imagers, much remains to be improved about their ability to accurately capture colour [3]. The role of colour discrimination falls to filters, and it is these that must be improved in order to improve colour capture in colour digital imaging systems. The process of colour reproduction of an imaged scene involves estimating the device-independent colour information of the scene from the sensor's device-dependent electrical signals. This enables storage and reproduction of the colour information. Commercial cameras today incorporate three filtering channels; this is based on the same principles as the three colour-sensitive photoreceptors of the human visual system [4, 5]. The spectral response of the filters currently used in standard imagers is different to the human visual system, so a non-linear transformation must be applied to digital images so as to reproduce colours as they are perceived by the human eye. The non-linear transformation strongly depends on the particular camera/sensor and this is why the colour performance of digital imaging products varies so much. More uniform and accurate colour reproduction would only be possible if the sensors more closely mimicked the spectral response of the human eye.

Some applications demand more accurate colour information, or in some cases more specific spectral information, than can be achieved with three channel systems. Such applications tend to fall outside the consumer market, and lie in areas such as art conservation (where exact matches or replication of the original paint is sought) [6, 7, 8], medical diagnostics [9] and computer vision (where accurate colour information assists in detecting objects) [10, 11, 12]. For such applications multispectral imaging is required. Multispectral systems have a greater number of colour channels, each channel sampling a narrower band filter. By increasing the number of channels, such systems can obtain more spectral information, providing improved colour reproduction, or even complete spectral reconstruction, facilitating a number of more specialised applications. Benefits developed to suit these most demanding applications ought also to be valuable in the consumer market, so long as they are low cost and practical to implement using mass manufacturing techniques. Ideally the filters would be fabricated directly on the image sensor, as they are currently in standard trichromatic systems.

In contrast to the significant developments made in reduction of pixel size and increasing CMOS sensor functionality, standard imagers still use filters fabricated

using spin cast pigmented or dyed films [13, 14, 15]. The colour or spectral reproduction of which is limited by the dye or pigment properties, or the need to incorporate additional external filter systems if narrowband filters are desired. Although such technology has provided sufficient reproduction capabilities to date, there is scope for better integration within the CMOS process and to improve the colour reproduction performance of the current CMOS imager filtering technology. Such developments would aid in the continued drive to improve, and integrate colour imaging systems within emerging applications.

This Thesis describes the design, fabrication and characterisation of new filtering techniques that can take advantage of the latest technology advancements to provide alternatives to the current on-chip filter technology, with the aim to provide improved colour reproduction and compact multispectral imaging. The filtering solutions are developed whilst bearing in mind imaging industry demands for miniaturisation, reduced cost and improved performance. This Thesis will explore the application of plasmonic filtering for standard three-channel digital CMOS imaging systems. The application of plasmonic filters, in addition to helping to meet the future needs of imagers, also provides the ability to manipulate the filtering response to improve colour reproduction performance. In addition, due to the increasing demand for compact multispectral imagers, this work shall also explore the development of a multispectral filter set that can be incorporated within a CMOS imager. This would open the CMOS imaging market to many new applications. Such filter sets are expected to meet the future cost and size requirements of CMOS imaging, while improving functionality and performance. In order to make the incorporation of such filters a viable process, it is essential that any proposed devices can be integrated within future CMOS imager technology; all the filtering solutions proposed here shall be fully compatible with the CMOS process.

1.2 Scope of this Thesis

The aim of this Thesis is to look specifically at the advancement opportunities available for visible band filtering technology for use within the CMOS process. It shall develop filter sets based on two technologies, plasmonic and Fabry-Pérot

filtering, with the aim of applying these technologies to achieve on-chip CMOS imager filtering solutions. The filter sets are designed to enable improved ease of fabrication and integration in-line with the continued miniaturisation of imagers and CMOS production processes. In addition to this, it is hoped to explore the prospects of such technologies to achieve improved colour reproduction from that available in current on-chip filter technology. The specific aims of this work include:

- to develop an automatic procedure for the design of plasmonic filters with arbitrary spectral responses
- to apply the automatic design procedure to replicate the human visual response in plasmonic filters
- to develop an on-chip narrowband filter set for multispectral imaging
- to experimentally characterise the filters' spectral responses
- to quantify the improvement in colour reproduction that is afforded by these filters.

1.3 Thesis Outline

This section outlines the organisation and contents of this Thesis. The work in this Thesis (Chapter 4-6) is presented in general order in which it was performed.

Chapter 2 provides an introduction to colour imaging systems and theory. A theoretical framework for colour reproduction is presented and the procedure is outlined for converting the electrical signals of an imager in to a device-independent representation of the colour that can be applied in colour reproduction. The concept, and calculation, of colour error is also explored. This enables the visual error in colour reproduction to be quantified. An overview of the latest trends in the CMOS process and materials, and current CMOS filtering technology, are also provided.

Chapter 3 provides a review of the relevant simulation and measurement techniques. The finite-difference time-domain method is used to simulate the spectral

transmission of the filters; an introduction is given to the specific application of the two available software packages (TEMPEST and Lumerical). The genetic algorithm used in the automatic design procedure is outlined. A description of the T-Matrix analytical technique and spectrometry, as applied in the analysis, is provided.

Chapter 4 introduces the theory of surface plasmon polarations, plasmonic structures, and the application of such structures for filtering. A complete automated design procedure is presented for plasmonic filter devices. The resulting designs and the simulated performance, is shown. The procedure for fabrication is detailed and, finally, the filter measurement results are discussed.

Chapter 5 introduces the Fabry-Pérot filter and details the related theory. In addition an introduction is provided to effective medium theory. These concepts are combined to create a narrowband filter. Two proposed filter designs are presented. A numerical analysis validates the user of effective medium theory to describe the optical properties of the nanostructured cavities. The simulation, fabrication and measurement of the two filter designs is described.

In Chapter 6, the measured filter performance is analysed in a numerical image reproduction process. The colour reproduction performance is compared to an example CMOS imaging filter set. A genetic algorithm is applied to identify sets of multispectral filters that minimise the colour error. An example application for the designed multispectral filter set is also outlined.

Chapter 7 summarises the results and suggests further work.

Chapter 2

Background

This Chapter provides an overview of the colour imaging process; from describing colour in abstract terms, to capturing it with digital imaging systems. It will briefly describe relevant aspects of the development of imager technology, including production processes, fabrication technology and the materials used within CMOS imagers. This Chapter will also provide an overview of the range of filtering technology currently utilised and proposed in literature.

2.1 Imaging Systems

An imaging system records the spectral reflection (intensity as a function of wavelength (λ)) from a scene as depicted in Figure 2.1. This reflection spectrum is a combination of the illumination ($L(\lambda)$) and the reflection properties of the scene ($S(\lambda)$). For a traditional camera, a snapshot of the spectral reflection from the scene is obtained by opening and closing the camera shutter for a period of time known as the exposure time. For a traditional film-based camera, the scene is focused onto a light-sensitive film located at the focal plane of the camera's optics. Digital cameras are of similar construction, except the film is replaced with an electronic photosensor. Only digital imaging systems shall be considered in this work because digital imagers now dominate the market, and the filters described in this Thesis do not apply to analogue cameras.

For some applications it is the only identification of the colour content of

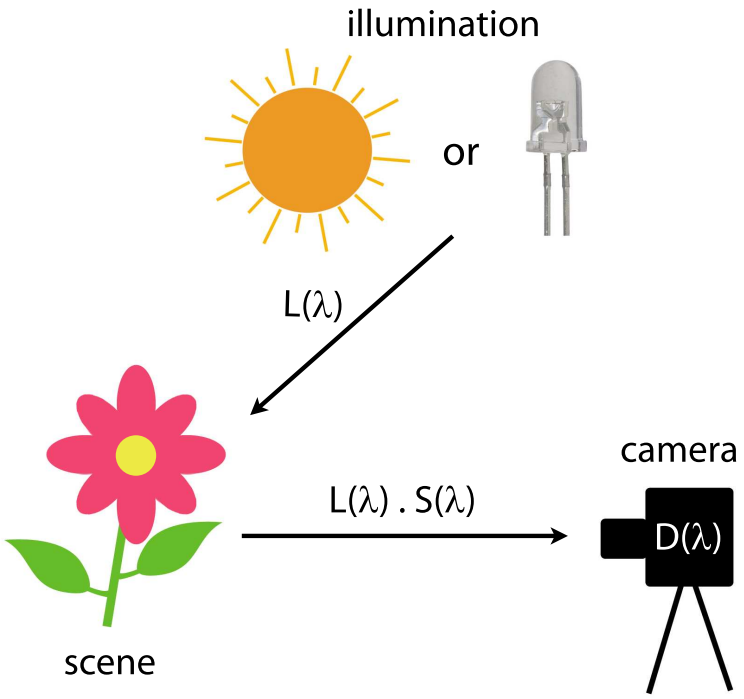


Figure 2.1: Schematic of the capture of the reflection from a scene.

the scene that is required to be extracted from the spectral reflection, and this can be achieved with trichromatic imaging. Other applications require a higher spectral resolution; for example to extract information about a specific wavelength band or for full spectral reconstruction. For such high resolution applications, multispectral imaging is used. The following section describes how such colour or spectral information is recorded and reproduced.

2.2 Colour Theory

The human visual system detects colour via the retina at the back of the eyeball. The retina contains two kinds of photoreceptor cells: cones and rods [5]. For this work only vision at normal daylight levels of illumination (photopic), as conducted by the cones, is considered. Cones come in three varieties that differ according to which part of the visible spectrum they are sensitive to; they are known as either the long, medium and short (L,M,S) wavelength cones. Colour

vision is achieved by comparison of the relative responses of each of the types of cones. The human visual system is not a wavelength detector (which would provide spectral information) but instead acts as an approximate linear integrator that can identify the colours contained in the scene. The integration approach enables a wide range of colour reproduction devices. A consequence of this system is that stimuli with different spectral power distributions can sometimes result in identical cone responses; this is known as *metamerism*. Trichromatic imaging attempts to extract and record colour information using three filter channels, each sampling different parts of the visible spectrum, as in the human visual system. A system that attempts to record and reproduce colour exactly as seen by the human vision system is known as a *colourimetric* imager.

2.2.1 Colourimetric Imaging

In 1931, the Commission Internationale de l'Eclairage[1] (CIE) developed a standard system for the specification of colour stimuli. They averaged data from two experiments with human participants, to create a set of values that indicate the relative amount of each of a set of the primaries required to create a colour match [4, 5]. These are known as tristimulus values.

The CIE set out to create the functions to be positive in order to enable simple calculation, and in addition they desired one of the functions ($\bar{y}(\lambda)$) to be the photopic luminance response, or the weighted sum of the sensitivities of the human eye. The result of this was the standard 2° observer colour matching functions (CMF); $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ as shown in Figure 2.2. The 2° observer CMF were derived using only the responses from cones (they attempted to eliminate any contribution from rods) and are intended to predict the response to small stimuli (for example, a thumbnail at arms length). These CMF are statistical approximations to represent the colour matching results of the average human observer. Since most images are composed of small patches of colours, this is an appropriate choice of observer system for most colour analysis problems.

In order to provide a numerical representation of the colours under the spec-

[1] the Commission Internationale de l'Eclairage (the french title for the International Commission on Illumination) was established in 1931 to provide a forum for discussion of, and develop standards relating to, the field of light, lighting and digital imaging

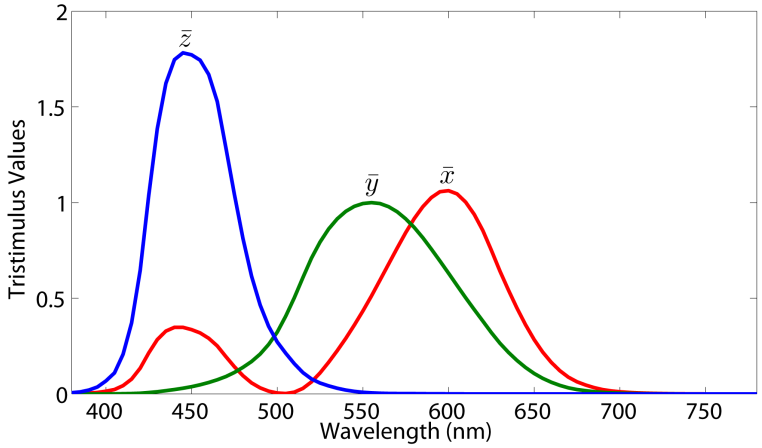


Figure 2.2: The CIE (1931) colour matching functions (CMF) for a 2° observer.

ified viewing conditions, a linear combination of these spectral sensitivity curves is used to map the spectra to the CIEXYZ tristimulus coordinate system values. For a surface with reflectance spectrum $S(\lambda)$ and illumination of relative spectral power $L(\lambda)$, the CIEXYZ values can be calculated by

$$\begin{aligned} X =& k \int_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{x}(\lambda)d\lambda \cong k \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{x}(\lambda)\Delta\lambda, \\ Y =& k \int_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{y}(\lambda)d\lambda \cong k \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{y}(\lambda)\Delta\lambda, \\ Z =& k \int_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{z}(\lambda)d\lambda \cong k \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)\bar{z}(\lambda)\Delta\lambda, \\ k =& \frac{100}{\int \bar{y}(\lambda)S(\lambda)} \cong \frac{100}{\sum_{\lambda_{\min}}^{\lambda_{\max}} \bar{y}(\lambda)S(\lambda)\Delta\lambda}. \end{aligned} \tag{2.1}$$

Figure 2.3 shows an example spectrum for an illuminant, the reflection spectra of a green and magenta surface and an example of a set of image system filter responses. The illuminant is D65 which is the CIE standard for typical northern sky daylight (and computer monitors). For printed paper, the standard illuminant is D50. The normalisation factor k is designed to assure that for a perfect white surface ($S(\lambda) = 1$) Y will be 100, irrespective of the illuminant. This normalisation imitates the adaptation to the illuminant that occurs in human

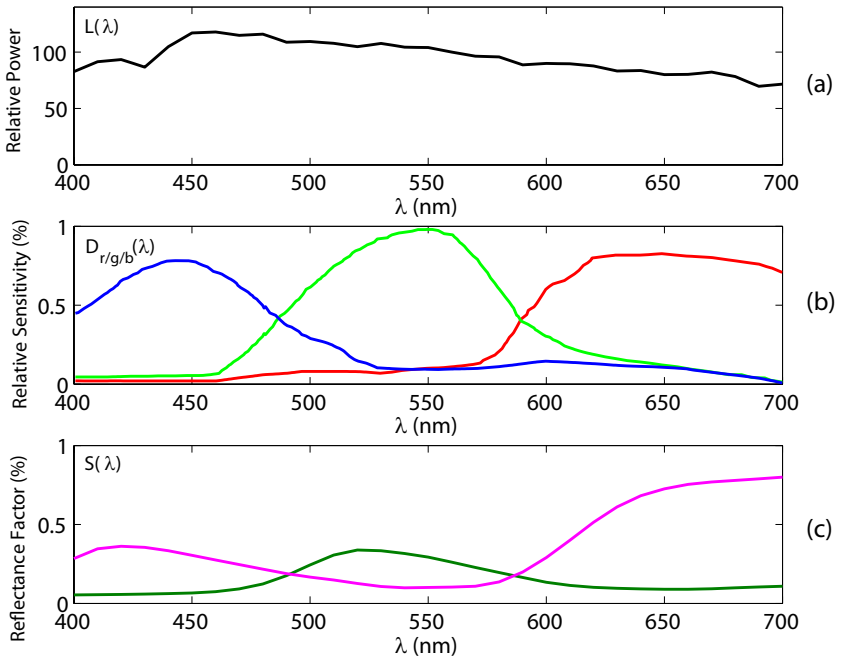


Figure 2.3: Example spectra (a) a plot of illuminant D65 ($L(\lambda)$), (b) a set of example red green and blue filter transmission spectra ($D_{r/g/b}(\lambda)$) [16] and (c) the reflectance spectra corresponding to the green and magenta patches of the GretagMacbeth ColorChecker chart.

vision (colour constancy). The illuminant cannot be deduced from the CIEXYZ components, so the illuminant data must be recorded separately to use in later processing steps. It is only necessary to know the relative spectral power distribution of the illumination. For most practical applications summations over 10 nm intervals, for the range $400 - 750$ nm, is sufficient for the calculation of the CIEXYZ values because the cone response curve is smooth. This is also the typical resolution and range of visible range spectrometer tools, hence the typical resolution of measured spectral data.

The CIEXYZ values provide a numerical description of a colour, but these are difficult to relate to the perceptual colour attributes. An attempt to overcome this yielded the development of chromaticity coordinates. Chromaticity coordinates aim to quantify the colour in terms such as its 'colourfulness', separate from brightness or luminance. The chromaticity coordinates of a colour can be

calculated from the tristimulus values using

$$x = \frac{X}{X+Y+Z} \qquad y = \frac{Y}{X+Y+Z} \qquad z = \frac{Z}{X+Y+Z} = 1 - x - y. \quad (2.2)$$

The value of z can be deduced from x and y, so it is convention to quote just x and y in addition to the Y tristimulus value. All colours that differ only by luminance (i.e. they have identical chromaticities), map to the same point on the chromaticity diagram; quoting Y retains the luminance information. The chromaticity diagram is shown in Figure 2.4. Contained within the area enclosed by the solid line (the spectral locus) are all the chromaticities visible to the human eye (gamut of human vision). The gamut of the sRGB colour space is also plotted (dashed line). The set of colours that can be reproduced with an sRGB system is significantly reduced from the gamut of human vision.

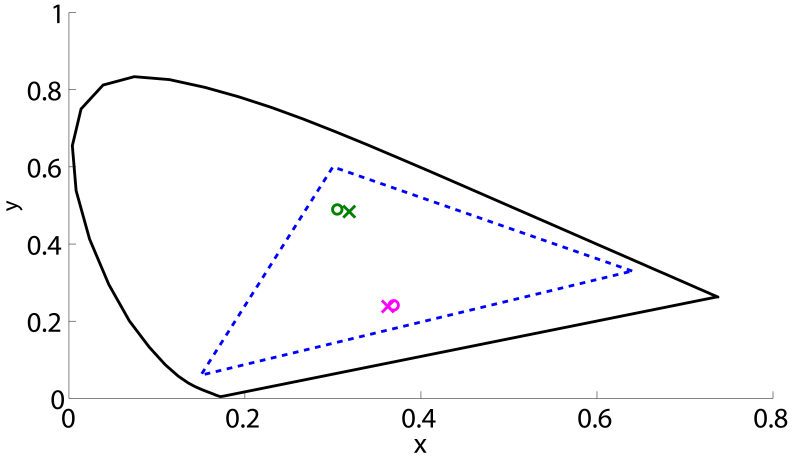


Figure 2.4: Spectral locus plot of the colour matching functions (−) and sRGB colour space (−−) using illuminant D65 [17]. The two points plotted correspond to the xy coordinates of the spectra ($S(\lambda)$) shown in Figure 2.3(c). These are calculated using the CIEXYZ values derived using Equation 2.1 (∘), and the XYZ values derived from Equation 2.3, using $D_{r/g/b}(\lambda)$ from Figure 2.3 and an appropriate transform matrix (×).

The colour space of a colour reproduction device such as a monitor or printer is defined by the available ink colours or phosphors. A similar limitation applies to colour imagers because of the colour filters - it may not be possible for a given image sensor to uniquely detect every colour in a standard colour space. To map

colour directly between such a variety of different device dependent detection and reproduction colour spaces would be impractical. The CIE system therefore is applied as an industry standard; to enable the device independent communication and storage of colour data. This colour space is the basis for all colour management systems.

Using the CIE system as a standard for communication of colour, the capability of an imaging device to accurately reproduce colour is determined by its ability to accurately transform the system signals to the tristimulus values. Every conversion to and from a device colour space to the CIE system will introduce error due to the variation of the filter spectra from the CIE functions. This is shown in Figure 2.4 where the xy coordinates (from a system with filter function $D_{r/g/b}(\lambda)$) differ slightly from that using the CIEXYZ values for the same patch spectra. To reduce the error introduced in conversion, the filter spectra should replicate the CIE function; the use of such filters will enable colourimetric imaging.

2.3 Colour Characterisation of an Imaging System

Each channel in an imaging system incorporates a filter in front of to the photodetector to enable the capture of the colour information in the scene. A trichromatic imager has three channels, akin to the human visual system. These are referred to as the red, green and blue channels, due the spectral transmission of the filters. To characterise an imaging system is to transform the device-dependent signals (r, g and b) into device independent colour data (CIEXYZ). The trichromatic

response of a channel in the image system is given by

$$r = \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)D_r(\lambda),$$
$$g = \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)D_g(\lambda),$$
$$b = \sum_{\lambda_{\min}}^{\lambda_{\max}} L(\lambda)S(\lambda)D_b(\lambda). \tag{2.3}$$

To calculate Equation 2.3, the full system response ($D_{r/g/b}(\lambda)$) must be known for each wavelength. To measure the spectral sensitivity of the imaging system, a monochromator and a radiance meter is required. An alternative method of determining the spectral response of the system, without the need for such specialist equipment, is generally applied for camera systems. The system captures images of patches of colour with known CIEXYZ values. The measured r, g, b signals obtained from the image enables the development of a mapping procedure, and therefore characterisation, of the imaging system. This is a cheaper and easier method however the sensor response is only valid for the particular illuminant.

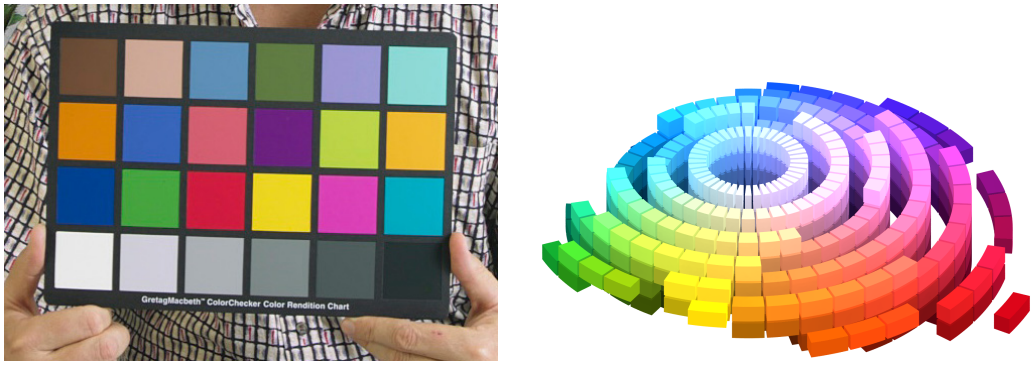
In this work only a comparison of the filter performance is required, therefore the complete imager response is not included. The filters' spectral performance is obtained using a spectrometer, and in this case D_r, D_g, D_b define the spectral response of the red, green and blue filters respectively. If $D_{r/g/b}(\lambda)$ is known, Equation 2.3 can be used to obtain the r, g, b response, and Equation 2.1 for the corresponding CIEXYZ values, for any illuminant and sample spectra.

2.3.1 Targets

To obtain mapping between a sensor output and a colour space, coloured patches with a known spectral response, and therefore a known point in the colour space, are required. Measurement of every possible colour to be imaged in future is impractical. The measurement of a few colour targets, covering the range of interest, enables the estimation of the device response to all future (as yet unmeasured) colours.

Such a characterisation of an imaging system uses a set of colour samples known as a 'training set'. In addition, another 'test set' may be required to quantify the accuracy of the conversion of the sensor signals to the corresponding tristimulus values or spectra. For a general imaging system the training patches should cover the entire colour space. For more specialised applications a colour patch set covering the range of interest can be used. The imager therefore will only be 'trained' for that section of the colour space.

There is no one universal colour sample set for the characterisation of an imaging system but there are several standard sets that are widely used. These include the Gretag-Macbeth ColorChecker Color Rendition chart (CCCR) [18], IT8.7 target for scanner calibration [19] and Munsell Book of Color [20]. Other colour sample sets, that are specific to the application, can be composed from reflectances that are to be reconstructed; for example colour samples of painting pigments or natural objects. Alternatively a set can be composed of samples selected using algebraic methods [21].



(a) (b)

Figure 2.5: Standard colour sets used: (a) GretagMacbeth ColorChecker Color Rendition chart and (b) a sRGB estimation of the Munsell Book of Color patches (reproduced from [22]).

As the characterisation conducted in the context of this work does not involve the requirement to manually image each sample, then the number of samples does not need to be limited. In Chapter 6 the CCCR chart shall be used, which has a set of spectral reflectances mimicking that of natural objects, such as foliage and human skin, and the full Munsell set (Figure 2.5). These provide good coverage

of the desired colour space and are often used for general imaging applications such as those considered in this work. In addition the spectral information for each patch, as required for this analysis, is freely accessible [23, 24].

2.3.2 Transformations

To obtain device independent colour information from an imaging system, and to quantify this and translate it for output, various transformations are required. For this work, the first transformation required is to convert the system output values to the tristimulus values. The tristimulus values can then be transformed for the reproduction of the scene on another device, or measurement of colour error. This process is described in the next section.

2.3.2.1 r, g, b to XYZ

There are several ways to map between the sensor output and the CIE colour space. These include look-up tables with interpolation, (least-squares) polynomial modelling, neural networks, and fitting data to models of the imaging system. In this work, a transformation matrix is formulated using a polynomial and a least-squares fitting technique. This required less computational and memory overhead and therefore is also suitable for in-camera processing. It is also more straightforward to compute compared to the time consuming training process required in neural networks. For N colour samples, the corresponding r, g, b imager response is represented by vector ρ_i, where $(i = 1...N)$. The corresponding XYZ tristimulus values can similarly be represented by a 1×3 vector x_i. The transformation between these vectors can be achieved by a linear transform ($\rho_i = [r\ g\ b]$), however improved mapping can be achieved by use of a polynomial, where the vector is expanded using more terms. For example a first order polynomial,

$$\rho_i = [rgb\ rg\ rb\ gb\ r\ g\ b\ 1] \tag{2.4}$$

or a second order,

$$\rho_i = [r^2\ g^2\ b^2\ rgb\ rg\ rb\ gb\ r\ g\ b\ 1], \tag{2.5}$$

are just two of the various combinations that can be applied. The number of terms is limited by the desired accuracy, number of colour samples and the computational intensity. In this work, Equation 2.5 is applied.

If $\mathbf{R}$ is defined as the matrix of ρ_i , where $\mathbf{R}$ is a $i \times 8$ matrix if Equation 2.4 is used, or a $i \times 11$ matrix if using Equation 2.5. The mapping from r, g, b to XYZ can be defined by

$$\mathbf{H} = \mathbf{AR}, \tag{2.6}$$

where $\mathbf{A}$ is the unknown transformation matrix of coefficients, which aims to minimise the error across all the colour samples, and $\mathbf{H}$ is a matrix containing the CIEXYZ values for each colour sample.

If a set of patches of known CIEXYZ are measured, and the r, g, b values obtained, from this the transformation matrix can be derived [25]. The matrix of coefficients can be determined from $\mathbf{H}$ and the pseudoinverse of $\mathbf{R}$, denoted $\mathbf{R}^+$:

$$\mathbf{A} = \mathbf{R}^+\mathbf{H}. \tag{2.7}$$

Mapping of the r, g, b channels to the CIEXYZ colour space is achieved using Equation 2.8 which minimises the sum of the squared residuals.

$$E = \sum_{i=1}^{N} (x_i^T - A\rho_i^T)^2 \tag{2.8}$$

The T superscript symbol denotes the transpose of the matrix. A least-squares solution is

$$\mathbf{A} = \left(\mathbf{R}^T\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{H}. \tag{2.9}$$

Using $\mathbf{A}$, formulated from a set of training samples of known r, g, b and CIEXYZ, future colour measurements can be mapped to the device independent CIEXYZ colour space using Equation 2.6.

2.3.2.2 XYZ to CIELAB

The CIE system is not uniform in terms of the visual perception; steps of equal distance in the chromaticity coordinates vary in their perceived distance. This was partially overcome in 1976 with the development of the CIELAB system of

colour specification for surface colours. CIELAB is a 3D colour space where a^* and b^*, the redness-greenness and yellowness-blueness respectively, form a plane and the lightness axis L^* is orthogonal to this plane. This is a non-linear transform of the XYZ values which aims to create a more perceptually uniform colourspace. The CIELAB coordinates can be calculated as follows

$$\begin{aligned} L^* &= 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16, \qquad \text{if} \quad \frac{Y}{Y_n} > 0.008856, \\ L^* &= 903.3\left(\frac{Y}{Y_n}\right), \qquad \text{if} \quad \frac{Y}{Y_n} \leq 0.008856, \\ a^* &= 500\left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right], \\ b^* &= 200\left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right] \end{aligned} \tag{2.10}$$

where X_n, Y_n and Z_n are the illuminant coordinates. This is usually calculated for a perfectly reflecting surface and therefore is equivalent to the illuminant [26].

2.3.2.3 XYZ to sRGB

For perfect reproduction of the XYZ tristimulus values, for display on a monitor or other output device, it is required that the device reproduce the CMF. However there is no output device that can achieve this because they are limited by the phosphors or inks used. As a result, several standard RGB spaces have evolved, either to match the display system or for technological reasons. Many exist, including Adobe RGB and Apple RGB [27]. These are designed for various applications, for example the printing industry, and cover large sections of the visible colour space, but sometimes extend outside of it. A general purpose space is sRGB, the spectral locus for which is shown in Figure 2.4. This represents the North-American TV standard and is compatible with a large fraction of recorded broadcast media. If the display properties are unknown, sRGB is often assumed. For this reason, sRGB is applied in this work to reproduce any colour from the CIE standard for visualisation.

For sRGB, white is defined with reference to the D65 illuminant. If the XYZ values are defined under a different illuminant to that required, they can be adapted to the destination illumination by application of the Bradford transform [27]. The RGB values can then be calculated with the transform in Equation 2.11, but note that the coefficient will vary for other colour spaces.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{2.11}$$

The RGB coordinates should then be rescaled so the 'white' coordinates are $(1, 1, 1)$, and clipped so there are no results outside the range 0-1. Similarly to how CIELAB is applied to convert XYZ to values that can be better associated to what is perceived, RGB coordinates are scaled according to the non-linear perception of the eye. This is done using Equation 2.12. This assigns more RGB values to the lower luminance levels where the eye is more sensitive. This is referred to as gamma correction, where $R'G'B'$ are the gamma corrected RGB values:

$$R^{'} = \begin{cases} \text{round}(255((1 + \text{offset})R^{\gamma} - \text{offset})) & \text{where } 1 \geqslant R \geqslant \text{transition} \\ \text{round}(255 \times \text{slope} \times R) & \text{where } R \geqslant 0 \end{cases} \tag{2.12}$$

For sRGB the parameters are

$$\begin{aligned} \text{offset} &= 0.055 \\ \gamma &= 0.42 \\ \text{transition} &= 0.003 \\ \text{slope} &= 12.92 \end{aligned}$$

The details of the components are explained elsewhere [27]. Here only $R^{'}$ is shown as $G^{'}$ and $B^{'}$ are calculated similarly. Equation 2.12 assumes an 8-bit system (therefore a 255 multiplication factor), and includes rounding to the nearest integer as required in a digital system.

2.3.3 Colour Error

It is useful, in the comparison of a sample and test colour, to define the magnitude of the perceived error between the samples. Such colour difference calculations however are only practical if the resulting error can be associated to what the eye perceives. With the development of CIELAB it became possible to define the colour difference between two colours by calculating the Euclidean distance between the two points in the colour space,

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{2.13}$$

where

$$\begin{aligned} \Delta L^* &= L^*_T - L^*_S, \\ \Delta a^* &= a^*_T - a^*_S, \\ \Delta b^* &= b^*_T - b^*_S. \end{aligned}$$

The subscripts S and T refer to the standard and trial points respectively. For industrial applications, the tolerance is around one ΔE^*_{ab} unit. One unit of error corresponds to a colour difference that only 50% of observers would identify as different - in other words, they are barely different. CIELAB ΔE^*_{ab} is the CIE recommended difference formula for large colour differences, i.e. $\Delta E^*_{ab} > 5$.

The CIELAB however is still not perfectly uniform, more recent work has lead to the development of a better system for the prediction of colour difference. In 2001, the CIE published the CIEDE2000 method for quantifying small colour differences [28]. It has additional correction terms to improve the performance of the calculation, for example to compensate for blue or other colours close to the achromatic axis [17]. The error is calculated as follows

$$\begin{aligned} \Delta E_{00} = \Bigg[&\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 \\ &+ R_T \left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)\Bigg]^{1/2} \end{aligned} \tag{2.14}$$

where

$$\Delta L^{'} = L^{*}_{T} - L^{*}_{S},$$
$$\Delta C^{'} = C^{'}_{T} - C^{'}_{S},$$
$$\Delta H^{'} = 2\sqrt{C^{'}_{T}C^{'}_{S}}\sin\left(\frac{\Delta h^{'}}{2}\right)$$

$$h^{'} = \begin{cases} 0 & \text{if } b^{*} = a^{'} = 0 \\ \tan^{-1}\left(\frac{b^{'}}{a^{'}}\right) & \text{otherwise} \end{cases}$$

$$\Delta h^{'} = \begin{cases} 0 & C^{'}_{S}C^{'}_{T} = 0 \\ h^{'}_{T} - h^{'}_{S} & C^{'}_{S}C^{'}_{T} \neq 0;\ |h^{'}_{T} - h^{'}_{S}| \leq 180^{\circ} \\ (h^{'}_{T} - h^{'}_{S}) - 360^{\circ} & C^{'}_{S}C^{'}_{T} \neq 0;\ (h^{'}_{T} - h^{'}_{S}) > 180^{\circ} \\ (h^{'}_{T} - h^{'}_{S}) + 360^{\circ} & C^{'}_{S}C^{'}_{T} \neq 0;\ (h^{'}_{T} - h^{'}_{S}) < -180^{\circ}. \end{cases}$$

The terms S_L, S_C and S_H are given by

$$S_L = 1 + \frac{0.015(\bar{L}^{'} - 50)^2}{\sqrt{20 + (\bar{L}^{'} - 50)^2}},$$
$$S_C = 1 + 0.045\bar{C}^{'},$$
$$S_H = 1 + 0.015\bar{C}^{'}T,$$

where

$$a^{'} = (1 + G)a^{*},$$
$$C^{'} = \sqrt{a^{'2} + b^{*2}}.$$

The terms G an T are calculated using

$$
\begin{aligned}
G &= 0.5 - 0.5\sqrt{\frac{\bar{C}_{ab}^{*7}}{\bar{C}_{ab}^{*7} + 25^7}}, \\
T &= 1 - 0.17\cos(\bar{h}' - 30°) + 0.24\cos(2\bar{h}') \\
&\quad + 0.32\cos(3\bar{h}' + 6°) - 0.2\cos(4\bar{h}' - 63°), \\
C_{ab}^{*} &= \sqrt{(a^*)^2 + (b^*)^2},
\end{aligned}
$$

and the rotation term R_T is given by

$$
\begin{aligned}
R_T &= -\sin(2\Delta\theta)R_C, \\
R_C &= 2\sqrt{\frac{\bar{C}'^7}{\bar{C}'^7 + 25^7}}, \\
\Delta\theta &= 30\exp\left[-\left(\frac{\bar{h}' - 275°}{25}\right)^2\right].
\end{aligned}
$$

For terms such as $\bar{C}', \bar{C}_{ab}^{*}, \bar{L}'$ and $\bar{h}'$ the arithmetic mean of standard and trial values are used [28].

Although no specific interpretation of the values obtained with the CIE2000 metric are identified in literature, there are interpretations of the CIE94 error formula (the recommended colour difference metric prior to the development of the CIE2000), which provide a suitable guideline. For this metric, $\Delta E = 1$ corresponds to the detection threshold of human vision, a difference of $\Delta E = 2$ is only just discernible for patches immediately adjacent to each other, while $\Delta E = 5$ is easily discernable [26].

2.4 Imaging System

In order to understand how digital imaging systems can be improved it is important to understand their construction and the current state of the art. The history of the digital imaging system, and its sensor architecture is discussed extensively in literature [29, 30, 31]; here an overview is provided. A modern digital imaging

system consists of five main parts excluding the control electronics. A simplified imaging pipeline is depicted in Figure 2.6.

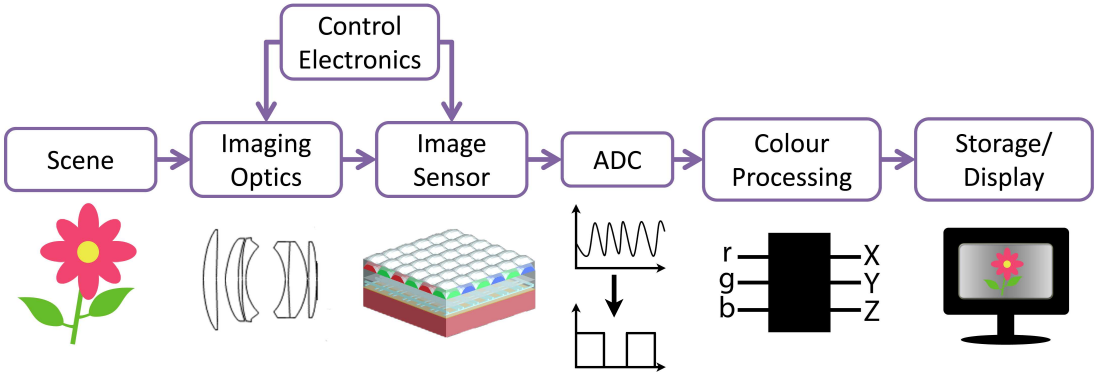


Figure 2.6: The imaging system pipeline.

The imaging process begins with the scene that is to be imaged. This work shall only examine the use of imagers for application in the visible light band; the section of the electromagnetic spectrum to which the human eye is sensitive. This range is roughly wavelengths (λ) of $390 - 780\,\text{nm}$ [32]. Although the scene may contain electromagnetic energy outwith this band, the imager aims only to record the contribution from within this band at each spatial point. The electromagnetic radiation incident on the imaging system, also referred to as the incident illumination, can be treated as electromagnetic waves or as a flow of photons (constituting discrete packages of energy). The amount of energy carried by a photon is determined by its wavelength [33]. A typical pixel records 10,000 photons per exposure. Psychophysical experiments show this is the level at which uniform photon noise becomes visible [34]. This can however be reduced with image processing and image noise removal.

The system's optics focus the scene onto a sensor [14]. Control electronics are required to control the focus and exposure. A cross-section of an example camera's optics and image sensor is shown in Figure 2.7. The image sensor consists of a two-dimensional array of pixels and appropriate circuitry for control and conversion of incident illumination to electrical signals. A photodetector converts the incident illumination from the scene into a photocurrent. Additional circuitry is used to read out this signal from the pixel array, converting this pho-

tocurrent to an analogue electrical signal. This analogue data is digitised using an analogue-to-digital converter (ADC) then processed, to provide colour information for storage or reproduction. Additional processing can improve imaging performance by adjusting the gain of the colour channels to account for the varying spectral performance of the system, and correction of defective pixels [16].

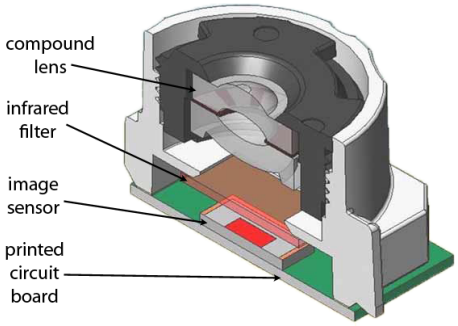


Figure 2.7: Cross-section of a camera module. From [35], labels added.

2.4.1 Image Sensor

This Thesis focuses on the development of filters that can be incorporated within the image sensor production process, so an understanding of the image processor architecture and fabrication is required to enable the design of filters that are compatible with the technology. There are two imaging sensor formats: charge-coupled devices (CCD) and CMOS imagers. The sensor technology and read-out architecture is different for each technology. The choice of CCD or CMOS is often made with reference to the system requirements, for noise, speed and power consumption. Both sensor formats convert the light to a charge at the pixel. This charge must be transferred to the charge detection amplifier, converted to a voltage and read out. Generally, in CCD, the charge is transferred across to the array boundaries and is read out sequentially. In CMOS the charge is not transferred; instead it is converted to a voltage by circuitry at the pixel. CCD has traditionally been the dominant image sensing technology as a consequence of historically having a smaller pixel size and superior noise and pixel-to-pixel variation performance. In CCDs, the common charge transfer components reduce the number of sources of pixel-to-pixel variation due to device mismatches, such

as matching of the individual amplifiers (known as fixed pattern noise). The ability to adapt the common amplifiers to minimise noise results in better noise performance from CCDs. CMOS sensors have individual amplifiers at each pixel and a higher dark current therefore in the past, have had significantly poorer noise performance than CCD. This however also provides the opportunity for faster readout speeds and low power operation [14]. CCD imagers often incorporate several additional chips to perform functions such as timing control and signal processing, as it is not economical to integrate these within the specialist CCD fabrication procedure.

The performance gap between the technologies has been reduced by design improvements to CMOS. CMOS sensors are fabricated using standard foundry processes, for this reason they can be fabricated alongside control circuits, timing circuits, analogue to digital converters (ADC) and signal processing on chip, resulting in single-chip cameras. CMOS sensors can take advantage of the drive for smaller dimensions in CMOS fabrication technology to reduce the complete system size and enable incorporation of additional signal processing and data handling on chip, providing greater functionality. Such additional functionality can be included to improve the noise performance. The increased, at sensor, circuitry however also reduces the area available for sensing of photons, therefore a balance must be maintained between the additional functionality and sensitivity.

Such technological advances, low power operation and the improved noise performance has resulted in the dominance of CMOS imagers in the low cost image market. Applications include webcams for portable computers and tablets, camera phones, automotive and astronomical cameras, optical mice and many more. In such applications, the advantages of the CMOS sensor far outweighs the reduction in noise performance of CCD. CCD still finds application in high-level imaging markets, such as scanners, copiers and video cameras, due to the low noise level and high quantum efficiency.

2.4.2 The CMOS Pixel

The light-sensitive region of the image sensing area of an image sensor consists of an array of pixels. Pixel dimensions of around $1.75 - 1.4\,\mu$m [36, 37, 38]

are common, however current CMOS technology permits sub-micron scale pixels to be fabricated [39]. The number of pixels in an imaging system varies with application and desired resolution. Recent commercial digital still cameras have had up to 16M pixels, while professional cameras have around 21M pixels. The latest smartphones have 8M pixels, however the Nokia 808 Pureview has a 41M pixel camera. A typical CMOS pixel consists of a polyimide micro-lens, filter, photodetector and an appropriate readout circuit. A cross-section of a pixel is shown in Figure 2.8. The lens focuses the light incident on the pixel surface onto the photosensitive section of the photodetector surface in the bottom layer of the stack, thereby improving the optical efficiency of the pixel (the fraction of photons on the pixel surface to that on the photodetector). The incident light is filtered such that only light from a specific component reaches the photodiode at that particular spatial point. The resulting electrical signal corresponds to the intensity of that specific colour at that point. This filtering step is required to enable reproduction of the colour information in later image processing steps.

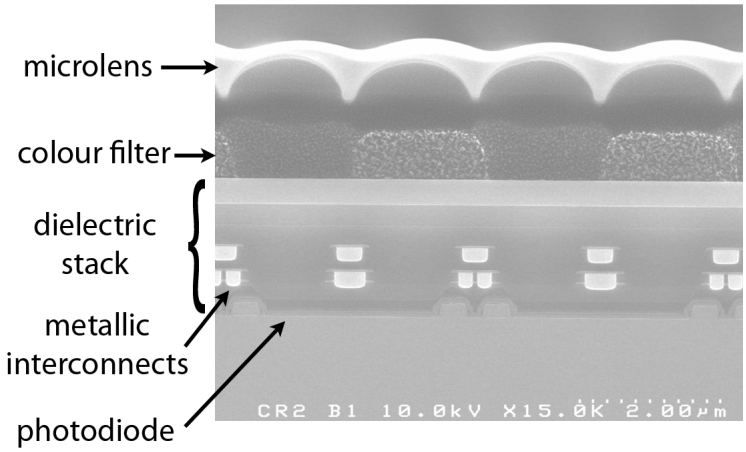


Figure 2.8: Scanning electron micrograph (SEM) image of the cross-section of a CMOS pixel [35].

If photons enter a semiconductor with energy greater than that of the bandgap energy then they are absorbed. Absorption occurs via a quantum interaction: an electron is excited from the valance to the conduction band, creating an electron-hole pair. If an electric field is applied, then the photon absorption results in a photocurrent. For silicon, the bandgap energy is 1.1 eV, and therefore can detect

wavelengths shorter 1.12 μm, for this reason it is often used as the substrate for visible wavelength detectors. Below the filter there are several metallic interconnect, dielectric spacer and passivation layers that compose the pixel circuitry, and a photodetector at the pixel floor. For CMOS imaging, the photodetector is generally composed of a reverse biased p-type - n-type (PN) semiconductor junction or a p-type - insulator - n-type (PIN) diode [40]. In CMOS imagers the photocurrent is integrated over a period of time and is read out after this exposure time. There are multiple circuit architectures available for the purpose of conversion from charge to voltage for read out. These include passive pixel sensor (PPS), the 3T active pixel sensor (APS), 4T APS (PIN), photogate APS (5T) and the digital pixel sensor (DPS) with an integrated ADC [29, 30, 33]. In general the advances in pixel architecture results in an increase in the number of transistors within the circuit, with the aim of reducing noise and increasing read out speed, at the expense of reduced fill factor (the ratio of photosensitive area to total pixel area). The highest fill factor is that achieved by the PPS design, which can be as high as 90%, however it has high readout noise. APS has reduced fill factors of between 50-80% [37, 41] but improved noise performance [33, 42]. The PIN APS noise performance is improved yet further by the ability to incorporate correlated double sampling, where the pixel is read twice (at the start and end of integration) enabling the removal of the reset and some fixed pattern noise [14, 29]. This particular development aided in increasing the popularity of CMOS sensors. The choice of architecture depends on the requirements of the specific application and the cost. The most common are the 3T and 4T architectures.

The challenge for the pixel circuit designers is to maintain sensitivity and quantum efficiency at the same time as reducing pixel size. The most recent developments include sharing of transistors between pixels, for example the 1.75T and 1.5T architecture, increasing the photodetector performance while minimising the reduction in fill factor [42]. There has also been an increased drive to look at how back-end of line (BEOL) processes can be adjusted to aid sensor performance, particularly for the latest generation of $\leq 1.75\,\mu$m pixels. Designers have started to design pixels with balanced interconnect patterning and dummy metal interconnects to improve optical confinement [37, 38]. Similarly light guides are being included in the pixel (Figure 2.9(b)) to channel the light down the pixel

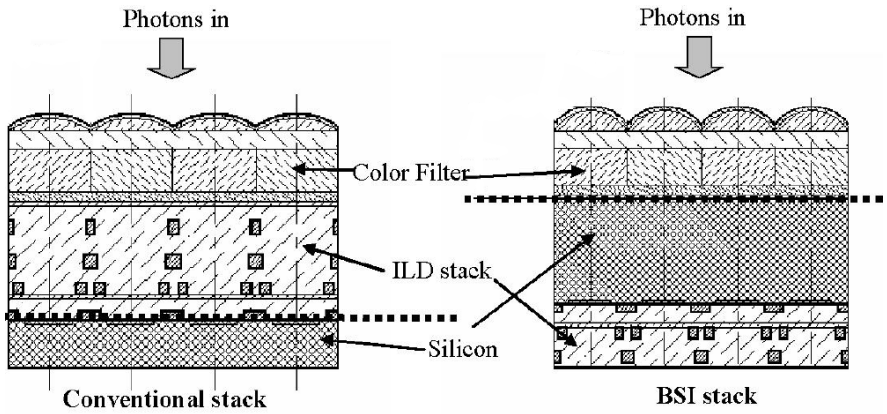
tunnel of a conventional stack and focus it on the photosensitive surface, reducing optical crosstalk [36, 43]. Manufacturers have also begun to develop backside illuminated (BSI) sensors, where the photosensor is developed on top of the interconnect stack, therefore theoretically providing 100% fill factor [44]. A schematic of this structure, and a traditional stack structure, is shown in Figure 2.9(a). BSI sensor production uses a polishing process to reduce the thickness of the photodiode. This technology was initially considered too costly for application in consumer electronics, and it also presented several design challenges such as the backside and frontside alignment and electrical crosstalk, however the production of the first generation of such imagers began in 2009 [37].

The resolution of an imaging system is limited not only by the minimum pixel dimensions but also by the system optics. The resolution of the optics is determined by the diffraction limit. This is shown in Figure 2.10 for an aperture, however the same principle applies to lenses whose boundaries define the aperture. For a circular aperture this causes a circular pattern of low secondary maxima around the main lobe. This is know as the Airy pattern and is shown in Figure 2.10(b). The diameter of the main lobe (d), which is considered to determine the minimum resolution, is given by

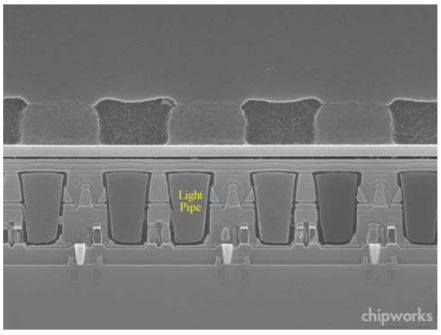
$$d = 2.44 \cdot \lambda \cdot f\#. \tag{2.15}$$

The main lobe contains about 84% of the incident light intersity [45].

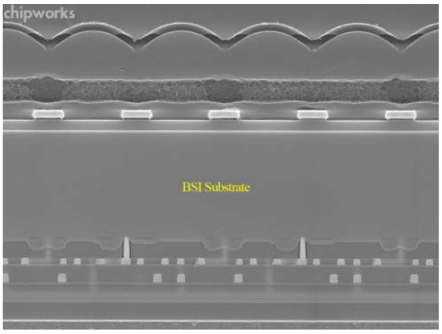
For a $f/2.8$ lens at $\lambda = 400\,\text{nm}$ (providing the minimum d), this corresponds to a spot size $2.73\,\mu\text{m}$. To avoid spatial aliasing, the main lobe of the point spread function must be sampled by two pixels. Smaller pixels will not improve resolution beyond that set by the choice of lens and wavelength. In terms of pixel width, for a $f/2.8$ lens, this means a $1.37\,\mu\text{m}$ wide pixel, where each pixel corresponds to a pixel in the resulting image. For a standard colour imager, each pixel in the resulting image is formed from the colour information provided by a 2×2 pixel array. A single pixel in the image therefore can be sampled twice, but each sample can contain the four pixel systems to provide colour information (as shown in Figure 2.11). So for a system such as this the minimum pixel size is $d/4$. Although in practice many lenses fail to meet the diffraction limit [46].



(a)



(b)



(c)

Figure 2.9: Schematics and scanning electron micrographs of imagers: (a) a comparison of a back side illuminated sensor and a standard CMOS sensor. The dashed line indicates the silicon interface for photon absorption [44]. Scanning electron micrograph images of (b) a standard CMOS pixel with lightpipes (microlenses not shown) and (c) a first generation back side illuminated sensor[37].

This often defines the maximum resolution of an imaging system, however there are other motivations aside from increased resolution that drive the continued reduction in the pixel size, for example greater toleration to pixel defects. It can also enable the development of other forms of imaging such as multispectral, 3D imaging or lightfield imaging.

This section introduced the CMOS pixel and provided an overview of the main technological developments within the field, as relating to improving the capabilities of the CMOS sensor. This drive for improvement has led to the increasing importance and emergence of CMOS as the dominant imaging technology. In

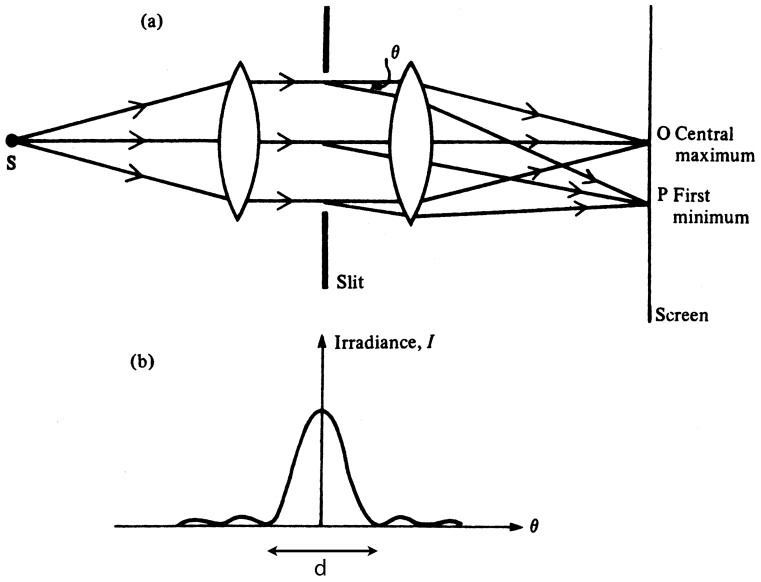


Figure 2.10: Demonstration of the effect of diffraction at the lens boundaries: (a) schematic of diffraction at an aperture (b) the resulting distribution of the incident light (Airy pattern)[45].

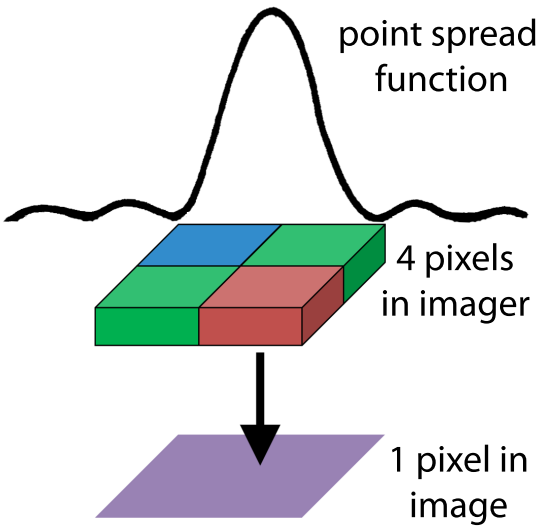


Figure 2.11: Illustration of the point spread function in relation to the pixel array.

the past, research efforts had focused mainly on the advancement of the pixel architecture and processing technology. It is only fairly recently that it has been considered possible to advance image sensors by improving the filter technology, with the added benefit of cost savings.

2.5 CMOS Processing and Materials

CMOS imager sensors are made using the same highly-integrated circuit technology and fabrication processes as microprocessors and memory chips. For this reason, the CMOS imager has benefited from the computer industry's demand for faster, lower-power, compact electronics. Similarly, benefits accrue from the investment, infrastructure, mature process technology and cost efficiencies that come with the large global production of CMOS devices. Due to the shared technological processes, CMOS imagers can be integrated with control electronics and other sophisticated processing circuits on a simple CMOS substrate. This provides a low cost route to highly integrated imaging products

The continued reduction in the feature size of the CMOS process results in CMOS imagers also migrating between technology nodes so that the overall chip size can be shrunk by scaling the control electronics, even if the pixels themselves are not scaled as per the discussion in Section 2.4.2. Details are provided in Table 2.1 of several nodes (achieved in the past and future predictions) and the critical dimensions that can be achieved, as detailed in the International Technological Roadmap for Semiconductors (ITRS). The technology node typically used to fabricate CMOS imagers, and the voltage requirements, lags behind the technology node used for computing because, often, the most advanced processes have not yet been optimised for imaging. The very latest $1.75, 1.4\,\mu$m and even sub-micron pixels are being fabricated in the $0.1 - 0.09\,\mu$m technology, however the design rules and front-end-of-line (FEOL) can date back to the $0.18\,\mu$m process [37, 39]. The CMOS process used does scale at the same pace as standard digital CMOS processes at a factor of 20 [42]. A plot of the reduction in pixel size and technology node used in design is shown in Figure 2.12. The technology node therefore gives insight to what is being achieved currently in logic technology, and therefore what can be expected in the future for imaging. The $0.13\,\mu$m process was broadly

Year	1999	2004	2010	2011	2012	2013	2014
DRAM 1/2 pitch (nm)	180	90	45	40.09	35.72	31.82	28.53
No. Metal Layers	6-7	10	12	12	12	13	13
Metal 1 pitch (nm)	500	214	102	76	64	54	48
Metal 1 A/R for Cu (Al)	1.4 (2)	1.7	1.8	1.8	1.8	1.9	1.9
Intermediate pitch (nm)	640	275	102	76	64	54	48
Intermediate A/R for Cu (Al)	2 (2.2)	1.7	1.8	1.8	1.9	1.9	1.9
Global pitch (nm)	1050	410	154				
Global A/R for Cu (Al)	2.2 (2)	2.1	2.34	2.43	2.34	2.34	2.34

Table 2.1: Table of semiconductor nodes and minimum microprocessor unit wiring dimensions [47, 48, 49, 50].

adopted by image sensor manufacturers around 2006, to coincide with the migration to the 2.2 μm pixel. Some companies merely adapted the depreciated DRAM fabrication process for colour image sensor production.

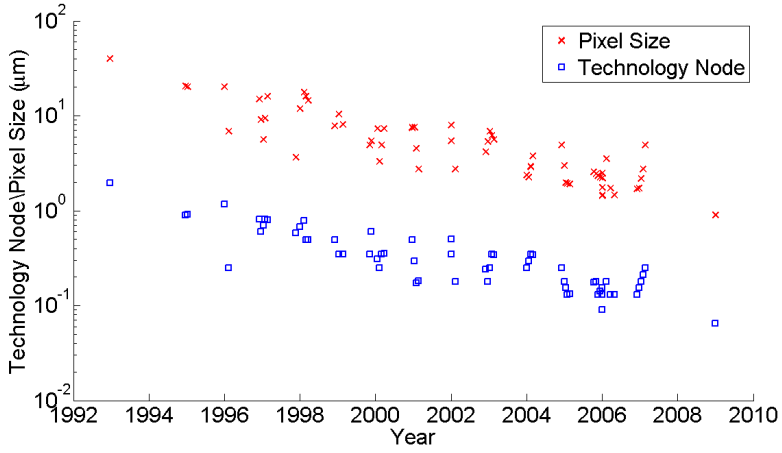


Figure 2.12: The evolution of the CMOS pixel size and the technology node used for fabrication. Data extracted from [42] (Figure 1) with additional data points from [37, 39, 51].

As detailed in Table 2.1, CMOS technology utilises an increasing number of layers, therefore increasing stack height, as the thickness of the interconnect layers scales at a rate slower than the planar dimensions. For a 0.18 μm process, the stack height can be around 10 μm. An increasing stack height is detrimental to

the operation of imagers because the filters and microlenses are placed further from the sensing surface. This results in reduced pixel performance at oblique angles and increased optical crosstalk (where light incident on a pixel is leaked into the neighbouring pixels), therefore colour separation becomes difficult. Imager production processes remove some layers to reduce this distance, however the stack height does increase with migration to the next technological node.

The continued scaling of CMOS technology aids the mass production of imagers as the accommodation of more chips on a wafer results in the reduced cost per chip. The scaling factor for cost is approximately p^2 where p is the pixel pitch (assuming the number of pixels is unchanged). The energy required to read the chip also scales at approximately p^2. However the pixel performance is also affected by scaling with a factor approximately p^{-1} with signal -to-noise and p^{-2} with dynamic range [42].

The fabrication of CMOS devices requires amongst other things multiple stages of deposition of oxides and nitrides using several variations of chemical vapour deposition; similarly metals are deposited via sputtering or evaporation processes. All such processes involved in forming the stack, interconnects, filter and microlens layers are collectively known as back-end-of-line (BEOL) technology. The most common metals used in the CMOS process today are aluminium (Al) and copper (Cu), although tungsten and other metals are also available. Silicon dioxide (SiO_2) and silicon nitride (Si_xN_y) are used as spacers or passivation layers. These materials are used for their electrical or dielectric properties, but they are also cheap and easy to process. Lithography and etching processes are used to form interconnects and other structures and chemical mechanical polishing is used to planarise surfaces. Ion implantation is used to create the p and n-type semiconductors (front end technology). Further details of such processes and their operational concepts can be found in literature [40, 52]. All of these are standard processes used for making various devices and are highly repeatable. The filters and microlenses however are formed on top of the stack in specialised BEOL processes.

Currently CMOS foundries use optical lithography. The minimum dimension that can be achieved with optical lithography is 40 nm half pitch, although smaller features can be achieved using techniques such as double patterning [53]. To meet

the demands of future technology nodes, manufacturers have begun to invest in the next generation of lithography tools. Extreme ultraviolet lithography [40] is the technology most have opted for, with several manufacturers purchasing production tools with a resolution of 22 nm [54]. If all the other infrastructure is ready, it is expected they will be in mass production by 2014 [53].

This development in the technological nodes, and therefore the lithography technique, will enable new devices with nanoscale features to be explored for application in colour filters, with a view to integration within the CMOS process itself, at future technology nodes. This would enable the capabilities of such infrastructure, when production ready, to be immediately exploited. This would aid in the continued development, both in the performance and fabrication procedure, of CMOS imaging. In addition to meeting the goal of producing mass market, high performance CMOS imagers, any suggested variations or additions to a foundry process must maintain compatibility with the CMOS design flow; i.e. they must continue to be fabricated where possible using standard fabrication processes and materials.

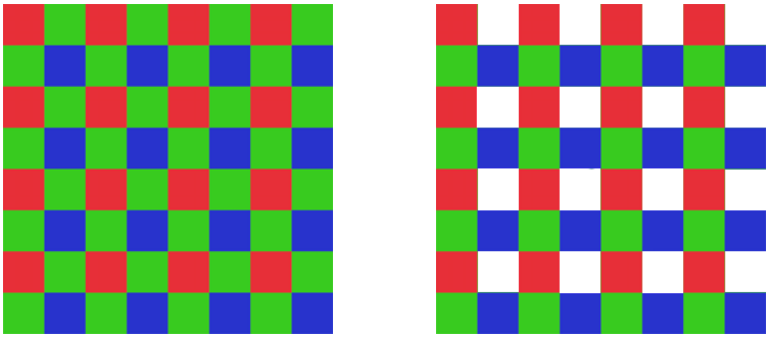
2.6 Colour Filtering

As indicated in Section 2.4.2, the photodetector in an imager is not wavelength selective within the visible band. To distinguish a section of the spectra for measurement and therefore enable colour imaging, the incident radiation on the sensor must be filtered prior to reaching the sensor surface. This results in an individual pixel being sensitive to only a selected wavelength band and thus the photodetector providing information to the intensity of the selected band at that point in space. The two imaging systems targeted in this work (multispectral and trichromatic) have different filtering requirements and therefore different technology to meet this. The next sections shall explore the various techniques applied to achieve the different types of imaging.

2.6.1 Trichromatic Filtering

The colour information of a scene can be captured by sampling three sections of the visible spectra at each point in the image, not unlike human vision. These spectral sections are usually referred to as the red (r), green (g), and blue (b) channels. Imaging systems employing three channels for colour reproduction are known as colourimetric, rgb or trichromatic imaging systems. The contribution of each of the channels, at each point in space, is usually obtained by the application of a mosaic of filters across the pixel array, known as a colour filter array (CFA).

Many patterns of CFA are available, of which the most common is the Bayer array [55]. The Bayer array is shown in Figure 2.13 consists of a repeated array of a 2×2 pattern of pixels, comprising one red, two green and one blue pixel. The green appears in the pattern twice because the human visual system has highest sensitivity around the middle wavelengths of the band. To reproduce the colour image, while maintaining spatial resolution, interpolation (or a demosaicing procedure) [14] is conducted at each individual pixel with contribution from the neighbouring pixels of different filter spectra.



(a) (b)

Figure 2.13: Colour filter array (CFA) (a) Bayer and (b) Honda *et al.*

Alternatives to the Bayer array were presented by Kodak [14] and Honda *et al* [56] in 2007. These patterns included un-filtered pixels (see Figure 2.13(b)). As the nature of a filter is to exclude significant portions of the spectra, this reduces the amount of incident light that reaches the sensor. If a non-filtered pixel is included then this can be used to derive the luminance with greater accuracy.

On exclusion of the non-filtered pixels the Kodak patten reverts to the Bayer array, therefore enabling conventional colour interpolation [14].

Currently the filters used within pixels use pigmented or dyed photoresist and are produced in an iterative spin process [13, 14, 15]. The resultant colour information from the sensor is device dependent as a consequence of the various different dyes used by different manufacturers. An example of a manufacture process is provided in [13], and [15] provides a list of example manufacturer patents in the reference section. The use of such filters involves specialist BEOL processes and materials. Their production is as complicated as the CMOS processes and add appreciably to the cost of the image sensor. Dye manufacturers must ensure the dyes are easy to apply, durable and resistant to humidity. This limits the dyes that can be applied and also the ability to tune the spectral response.

An additional issue with such filters is that as the pixel size shrinks, any variation in the dispersion of the pigment, and therefore the transmissive properties, will become more apparent in the system colour reproduction. The organic dyes also face issues of degradation due to light exposure [57] and are also transmissive in the near infrared (IR) region of the spectrum. As mentioned in Section 2.4.2 silicon has a bandgap of 1.12 eV, so a silicon-based sensor will detect any illumination of less than or equal to $1\,\mu$m. To remove any contribution from the unwanted IR radiation (here wavelengths between $0.75 - 1\,\mu$m) an additional IR cut-off filter is required, and this is often implemented within the imaging optics (Figure 2.7).

With the ever increasing CMOS stack height, efforts have been made to improve crosstalk issues by etching back some of the layers [51]. However the requirement for a protection layer between the CMOS stack and the filter, and the fact the filter itself cannot be reduced in height (currently $1\,\mu$m) as this will degrade the performance, does not aid the process. From these issues it can be seen that there are many opportunities in which the filtering process for CMOS imagers could be improved and therefore aid the miniaturisation, and reduce the processing cost of such systems.

2.6.1.1 Alternative Trichromatic Filters

To aid in the advancement of CMOS sensors, several alternative techniques to pigment or dye filtering have been proposed with the aim to offer better integration with the CMOS fabrication process and reduce the degradation of the filtering performance over time. One such alternative is the application of a photonic crystal structure [57]. A photonic crystal consist of a periodically structured material stack. Such structures can be designed so that light of a specified wavelength range, known as the photonic bandgap, cannot propagate through the stack. If a defect layer is introduced to disrupt the periodicity of the stack then a passband can be introduced within the bandgap. With careful design this can be used to create rgb filters. A photonic crystal design was presented by Inaba *et al* [57] which used dielectric layers, the total thickness of the stack was approximately $1.2 - 1.3\,\mu$m. The use of inorganic materials reduced the problem of degradation with time, however this is not a compact structure and the requirement for multiple deposition and lithography cycles to form the varying defect layer make such filters impractical.

Another filter structure was presented by Frey *et al* [58]. Frey's design utilised a multilayer Fabry-Pérot (FP) structure. The constructive and destructive interference between the metallic mirrors resulted in filtering of the incident light (see Section 5.2 for a more detailed description of FP filters). Although a much thinner structure (400 nm) than the photonic crystal filters, and that it does not require an IR filter, multiple deposition procedures and four lithography cycles are still required. In addition, silver was used for the mirrors which is expensive and not a material used in the CMOS process.

A complete alternative to transmissive filters was presented with the development of multiple stacked diodes [59] or variation of the electric field configuration at the depletion region of the diode [60]. Both these techniques take advantage of the wavelength dependent absorption of silicon; where the red, green and blue wavelengths are absorbed at different depths within the silicon. This can be exploited by placing diode junctions, or varying the electric field, at the appropriate depth for each channel absorption. Such designs can be implemented within the CMOS process, and have already been applied within a commercial colour im-

ager [59]. As each pixel provides the three colour channel information it reduces the need for interpolation between pixels and this can improve colour estimation. However, with the dimension of current imager pixels approaching the diffraction limit and therefore enabling multiple pixels per spatial point, this advantage may be negated. Both techniques required additional fabrication and circuitry at each pixel, thus increasing the size of the pixel from standard CMOS imagers. The filtering performance also has significant overlapping of response curves which will reduce colour reproduction performance, also an additional IR filter is still required.

Visible range filters have previously been demonstrated using metal-dielectric coupled nanostrip structures [61], these also however incorporate silver. Similarly another solution is the use of wire grid filters [62]. These are commonly applied in microwave and IR technology for wavelength filtering. With the improvement in lithographic technology it is now possible to fabricate such a structure on the scale required for visible light filtering (where the grating period is less than a wavelength). Such filters enable fabrication of the CFA in the standard CMOS process, removing the need for a separate CFA process and the cost associated with this. The filtering characteristics can be varied by altering the horizontal dimensions of the structure only, because of this, multiple filters can be defined in a single lithography cycle. Similar to the photonic crystal and FP filters, these can be incorporated on top of the chip stack or on top of the photodetector if BSI technology is used. The significant advantage of this technology is the use of standard CMOS materials and processing; with advancing technology such structures could be incorporated within the pixel stack. The implementation is shown in Figure 2.14, and results in lower pixel height and reduced colour crosstalk. The grating causes a polarisation dependence, that can be mitigated by adding adjacent filters with an orthogonal grating vector.

2.6.2 Multispectral Filtering

Multiple definitions of a multispectral imager exist, but here the term multispectral imaging is used in the context of imaging with four or more colour channels within the visible spectral range, as opposed to the aforementioned three channels

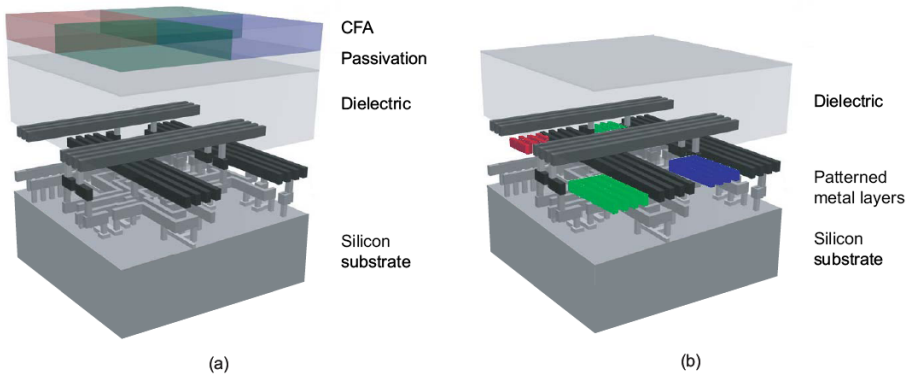
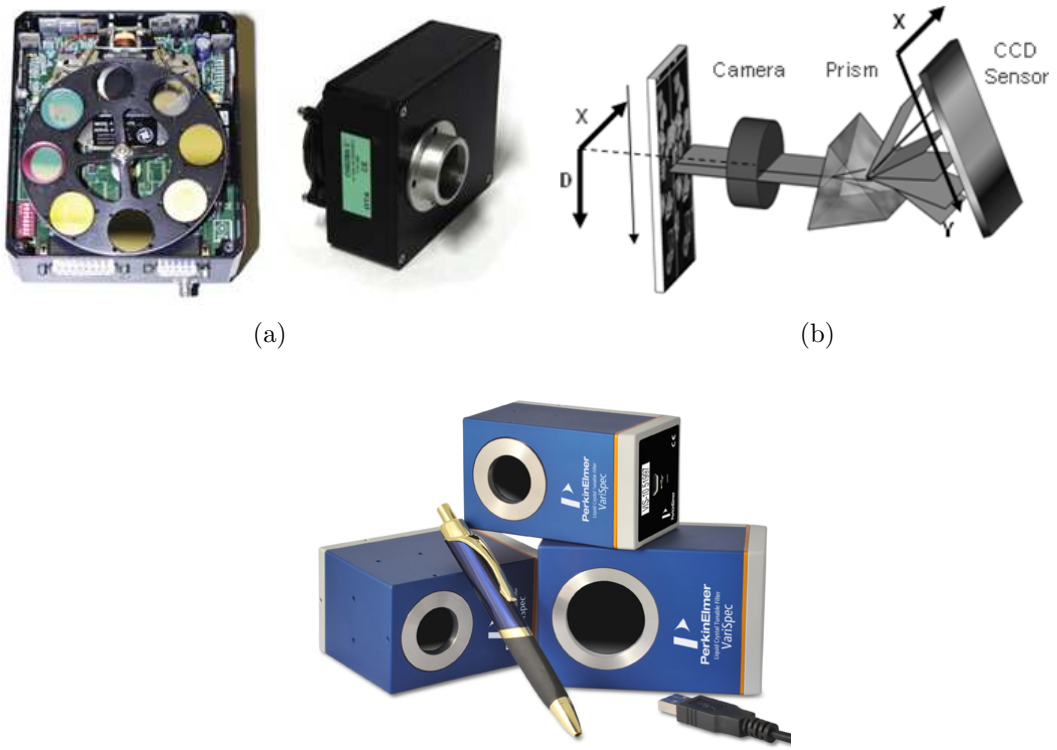


Figure 2.14: Demonstration of (a) a standard CFA and pixel stack, (b) the pixel stack with a nanostructured metallic filter within the pixel stack [46].

of trichromatic imaging. Traditionally, multispectral measurements have been conducted by a spectrometer using a beam-splitter technique. However with the advances of CMOS processing and an increase in the applications, such as agricultural monitoring systems [11, 63, 64], art conservation [6, 7, 8] and medical imaging [9], there is a demand for cheaper and more compact systems. There is no dominant method (like the dye based CFA for trichromatic imaging). Instead several technologies have been developed and each has its own advantages. The method applied is often dependent on the application. The possible approaches include beam splitting, sequential filtering and colour filter arrays.

2.6.2.1 Beam-splitting

Beam-splitting techniques split the beam of light in space with each spectral component focused onto a separate photodetector as shown in Figure 2.15(b). This is usually achieved with a prism or grating [65]. Such systems can obtain very narrowband selection, but require multiple exposures and careful alignment of the sensors with respect to the beam splitter, as well as line scanning because it uses a set of 1D arrays. This approach is only used in high-end applications due to the cost (for example, in spectrometers for scientific measurements where very high resolution is required).



(a) (b)

(c)

Figure 2.15: Example multispectral filtering systems (a) an imaging system using a filter wheel [66], (b) a schematic of a beam-splitting system using a prism [12] and (c) a VariSpec liquid crystal tunable filter [67].

2.6.2.2 Sequential Filtering

Sequential filtering involves swapping the filters in front of the imager, usually using a filter wheel as shown in Figure 2.15(a), with multiple successive exposures. The exposures are combined to form the final image. Using this method, the filters can be varied, and highly accurate (i.e. steep cut off) filters can be applied. This method provides high spatial resolution but takes more time; this requires that the object being imaged must remain still while the multiple exposures are completed. If variations occur between exposures, then colour-distortion will occur. There are also issues associated with the mechanical nature of the system (reliability, speed).

One sequential-filtering-based technology that has gained popularity is elec-

trically tunable filters. This technology has the advantage that it has no moving parts, unlike filter wheels. The most common tunable filter is the liquid crystal filter (as shown in Figure 2.15(c)), usually an adaptation of the Lyot filter [68]. Such systems use the electrically controlled birefringence of liquid crystals, to control the light as it passes through multiple filtering stages, resulting in the transmission of a narrowband of wavelengths at the end. Such systems consist of multiple layers of polarisers and liquid crystal material, this significantly increases the fabrication difficulty of such systems. In addition they require additional control electronics to perform the tuning and are polarisation sensitive. They also have the disadvantages associated with a sequential filtering system.

2.6.2.3 Colour Filter Arrays

Colour filter arrays, as detailed previously for trichromatic systems, involve the arrangement of multiple filters in a mosaic pattern across the pixel array. This enables an image to be created in a single exposure with reduced spatial resolution, overcoming a limitation of sequential filtering. It is easily integrated within the imaging system, however it does have issues associated with the spatial sampling of different colours at different positions and therefore will have spatial artifacts, particularly at colour boundaries.

Several possible approaches have been proposed for a multispectral CFA. These, similar to trichromatic filters, are based on FP or photonic crystal structures. Such filters are traditionally tuned by variation of the layer thickness. For multispectral imaging with an increased number of filters, this results in a significant number of lithography cycles, significantly increasing the fabrication demand. A solution to reduce the number of cavity lengths required is to alter the effective index of the structure materials. A system where gold particles were embedded within the cavity of a FP structure was presented by Mitra *et al* [69]. This would not be suitable for CMOS imaging because the process is uncontrolled, and gold is not compatible with the CMOS process (it can contaminate the silicon). Another solution was presented by Mehta *et al* [70], who etched air holes into a 41-layer photonic crystal. Two multispectral filters were demonstrated at wavelengths around 1500 nm, outside of the visible range. A simulation study has

demonstrated a similar filter structure at visible wavelengths, however these did not have a full visible bandgap and have not been verified experimentally [71]. These photonic crystal filters again consist of a thick structure with multiple layers, therefore the sensor will suffer from increased crosstalk.

The reduction in the pixel size, and the associated increase in resolution of imagers, has made multispectral imaging with a CFA a more attractive prospect. The key challenges to make that practical is finding a filter that can be fabricated in a minimal number of processing steps. The ability to incorporate such a multispectral CFA within the CMOS production process would enable the production of cheap, portable multispectral imagers.

Chapter 3

Methods

This Chapter introduces the techniques used to design and simulate the proposed trichromatic and multispectral colour filter arrays (CFA). The finite-difference time-domain technique was used for 3D simulations and coupled to a genetic algorithm for device optimisation. The 1D transfer matrix method is also described in addition to the measurement of the transmission of the fabricated filters using a microscope-integrated spectrometer.

3.1 Simulation Software

Several simulation methods were used to explore the performance of the devices presented in this Thesis. Each was appropriate for different applications, as will be detailed later. Here is an overview of each technique.

3.1.1 Finite-Difference Time-Domain (FDTD) Method

The finite-difference time-domain (FDTD) method models the propagation of light in a structure by numerically solving the time-dependent Maxwell's curl equations to second-order accuracy on a spatially discretised grid that defines the structure's material properties. The algorithm steps forward in small time increments, calculating the evolution of the electric (E) and magnetic fields (H) at each point, in an intuitively satisfying manner.

The FDTD method was first introduced by Yee [72]. He proposed the lattice

in which the domain is discretised into cells (with cell size Δx, Δy, Δz or if a cubic lattice cell size Δ), where the E and H components are staggered at half increments in space and time (where Δt is the time step) as shown in Figure 3.1.

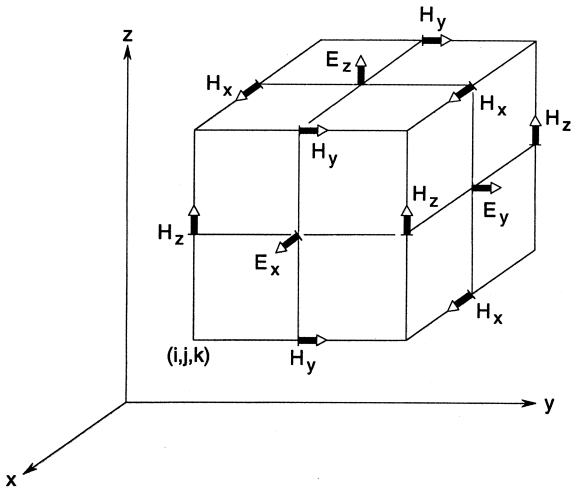


Figure 3.1: Position of the various field components within the Yee grid [73].

From Maxwell's time dependent equation of the electric field in partial differential form:

$$\frac{\partial \vec{E}}{\partial t} = \frac{1}{\varepsilon}\left[\nabla \times \vec{H} - \vec{J}\right] \tag{3.1}$$

it can be seen that a change of E with time results in a change of H in space. Similarly for H:

$$\frac{\partial \vec{H}}{\partial t} = -\frac{1}{\mu}\left[\nabla \times \vec{E} - \vec{M}\right], \tag{3.2}$$

were ε is the permittivity, and μ the permeability, of the medium. The electric and magnetic current density, $\vec{J}$ and $\vec{M}$ respectively, are given by

$$\vec{J} = \vec{J}_{\text{source}} + \sigma \vec{E}, \tag{3.3}$$

$$\vec{M} = \vec{M}_{\text{source}} + \sigma^* \vec{H}, \tag{3.4}$$

where σ is the electric conductivity of the medium and σ^* is the equivalent magnetic loss. Substituting Equations 3.3 and 3.4 into Equations 3.1 and 3.2, and

evaluating the vector components of the curl operators, yields a system of coupled equations which form the basis of the FDTD algorithm. For example the E_x component is defined as

$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left[\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - (J_{\text{source}_x} + \sigma E_x)\right]. \tag{3.5}$$

Using the rectangular coordinate system defined by Yee, a finite difference equation can be formulated from Maxwell's equations in discrete form [73]. For example, in order to derive the E_x component at time step $n + 1/2$, at a space point in the lattice $(i, j + 1/2, k + 1/2)$, which stands for the spatial location $(i\Delta x, j\Delta y, k\Delta z)$ where i, j and k are integers, is given by;

$$\begin{aligned}\frac{E_x|^{n+1/2}_{i,j+1/2,k+1/2} - E_x|^{n-1/2}_{i,j+1/2,k+1/2}}{\Delta t} = \frac{1}{\varepsilon_{i,j+1/2,k+1/2}}\cdot \\ \left[\frac{H_z|^n_{i,j+1,k+1/2} - H_z|^n_{i,j,k+1/2}}{\Delta y} - \frac{H_y|^n_{i,j+1/2,k+1} - H_y|^n_{i,j+1/2,k}}{\Delta z}\right. \\ \left. - J_{\text{source}_x}|^n_{i,j+1/2,k+1/2} - \sigma|_{i,j+1/2,k+1/2} E_x|^n_{i,j+1/2,k+1/2}\right].\end{aligned} \tag{3.6}$$

The coefficients are often fixed throughout a simulation, so they can be precalculated for improved efficiency.

$$\begin{aligned}E_x|^{n+1/2}_{i,j+1/2,k+1/2} = C_a(m) E_x|^{n-1/2}_{i,j+1/2,k+1/2} + C_b(m)(H_z|^n_{i,j+1,k+1/2} - H_z|^n_{i,j,k+1/2} \\ + H_y|^n_{i,j+1/2,k} - H_y|^n_{i,j+1/2,k+1} - J_{\text{source}_x}|^n_{i,j+1/2,k+1/2}\Delta),\end{aligned} \tag{3.7}$$

where $C_a(m)$ and $C_b(m)$ are constants associated with the medium as defined in [73]. From Equation 3.7 it can be identified that the electric field components at each point in space and time can be calculated from the previous time step's electric component, the neighbouring magnetic field components of the space lattice and the material properties. The equations are implemented in a leapfrog manner with the software calculating the field components at each point in space for a specified time, from previous stored data, then incrementing the time step

and repeating the process. Similar equations are derived for computation of the H field components. By placing a simulation 'monitor' at a certain location in the simulation mesh, the software records the field components at each time step and outputs them for analysis at the end of the simulation. By taking the discrete Fourier transform of the stored time data at the end of simulation the E and H field behaviour at each frequency, for that cell, can be determined. Using appropriate processing of the data, at the end of the simulation the device response to the incident illumination can be evaluated, including the transmission and reflection coefficients for a structure.

A restriction is placed on Δt and the cell size to ensure numerical stability; the field must not change significantly between cells therefore the cell size must be significantly less than the illumination wavelength. The minimum time step and/or cell size is determined by the Courant condition

$$S = \frac{c\Delta t}{\Delta}, \tag{3.8}$$

where S is the Courant number, and must be less than or equal to $1/\sqrt{no.\,dim}$ to ensure numerical stability i.e. less than or equal to $1/\sqrt{3}$ for a 3D simulation [73]. The cell size Δ is usually defined first to ensure fine enough resolution to define the structure geometry, and to ensure the propagating wavelengths can be resolved. The cell size is typically

$$\Delta \leq \frac{\lambda_{\min}}{10 n_{\max}}, \tag{3.9}$$

where $\lambda_{\min}$ is the shortest wavelength and $n_{\max}$ is the highest material refractive index present in the simulation domain. From this the minimum Δt can be calculated using Equation 3.8.

For examining the response of the device to a single frequency excitation, the incident light can be modelled within the simulation as a continuous monochromatic plane wave excitation. Alternatively a broadband Gaussian pulse can be used. A broadband pulse allows the response of the structure to be obtained for a range of wavelengths, in a single simulation.

Due to practical memory limits the simulation domain must be finite; bound-

ary conditions are required at the edge of the domain. For this work, three types of boundary condition are used: first, periodic boundaries, which treat the structure as if it is infinitely repeated. Second, perfectly matched layer (PML) boundaries which are a non-physical material that absorbs the energy reaching the edges of the simulation area. PML boundaries allow radiation to "exit" the simulation area without interfering with the fields inside. Third, Bloch boundaries, which are actually a form of periodic boundary and are used when a phase shift exists between each period of the structure. This is the case when a plane wave in the domain is excited at an angle off normal. For Bloch boundaries, accurate data can only be calculated for a single frequency of excitation per simulation.

In this work the FDTD method is employed to accurately determine the behaviour of the devices, and inform the choice of device dimensions such that the device performance is optimised as much as practicable before fabrication. The FDTD software available to this project includes the Lumerical and TEMPEST simulators. Both have the ability to simulate the desired structures at normal incidence, modelling both dielectrics and metals with various boundary conditions in both 2D and 3D. The reason two simulators are used is that TEMPEST is academic code without limitation on the number of copies that can be run simultaneously. It is useful for optimising device design. Lumerical has a simple to use user interface and additional features, not available in TEMPEST, such as Bloch boundary conditions. This therefore can be used for ease for individual simulations, or for studying device performance at non-normal angles of incidence.

3.1.1.1 Lumerical FDTD Solutions

Lumerical FDTD Solutions is a commercial software package that provides a graphical interface and a scripting language [74]. The scripting process however is not instinctive and relies on an understanding of the user interface, with the user required to identify the menu in which each file parameter is defined and to navigate between these. The software provides a tunable broadband source and the simulation data can be plotted using the inbuilt analysis tools. The analysis tool is useful for analysing individual devices because reflection and transmission data is automatically normalised to the source even for broadband pulses. For

more advanced analyses, e.g. comparing devices, the data is better exported. Then, analysis can be performed using a mathematical tool such as MATLAB or Octave. Lumerical incorporates several models for representing the dispersive behaviour of materials, however the material property behaviour can not be edited or improved beyond these models and the inbuilt fitting routine. The major advantage of this package is the implementation of Bloch boundary conditions, because this enabled simulation of angled planewave excitations. As this is a commercial software there was no access to the base code, and to run each simulation required access to a paid-for engine licence. This limited the access to the software such that it could not be used as part of the automated design procedure developed.

3.1.1.2 TEMPEST

TEMPEST software [75] is also a FDTD simulator. This software was initially developed by the Neurether Group at the University of California, Berkeley for application in the analysis of photolithography masks in the integrated circuit industry. It was adapted to enable 3D simulation by Wong [76] with further adaptations by Pistor to include PML boundaries and enable parallisation [77]. It was further adapted by Drysdale [78] to provide a broadband source. Although no graphical user interface is provided, thus making the software perhaps less accessible to a user of little experience, it had a major advantage in that the software is licensed without the need for an engine licence for each simulation, and with full access to the base code. This enabled easy combination of the simulator tool with the search algorithms as shall be discussed in Section 3.2. The search algorithm easily wrote input simulation files and could be batched and run automatically with the Unix prompt on a supercomputer without licencing issues. The ability to easily integrate the TEMPEST software with MATLAB or Octave software provided easy and efficient method of pre/post-processing. It did however lack the ability to model dispersive materials and so editing of the source code to include this capability was required for this work. This also provided the opportunity to implement a dispersive model with more control over the parameters than can be achieved using Lumerical. Further details on

the model and the implementation are provided in Appendix A. TEMPEST also did not have Bloch boundary conditions which limits the application to normal incidence simulations only (or angled simulations where the phase wraps around neatly). The simulation process was however much more transparent and enabled freedom of source code modification.

3.1.2 T-matrix

The transfer matrix (T-matrix) method is a quasi-analytical 1D approach [79]. It provides an alternative method for the simulation of some of the filter structures proposed. This method can be applied for the solution of wave equations incident on a 1D stack of homogeneous layers. The reflection or transmission at an interface between two media is described by Fresnel's equations [32]. For multiple interfaces, as present in the structures proposed, solving Fresnel's equations is tedious to do by hand. T-matrix is a straightforward technique that calculates the T-matrices at each layer interface, and these can then be multiplied together to provide the overall stack T-matrix. A phase adjustment is applied to each layer to model the layer's thickness. This can be used to quickly determine the transmission and reflection of a filter, if the effective material properties are known. It can be implemented as a script in MATLAB or Octave and is significantly less computationally demanding, and therefore faster, than the FDTD method (although it models less).

The T-matrix equations are developed as follows. Any wave of a given wavelength, with time independent form

$$E(z) = E_0(z)e^{ikz}, \tag{3.10}$$

incident at an interface between two media will experience partial transmission and reflection. Here E_0 is the maximum amplitude of the wave, k is the propagation constant and the position is z. If the media permittivity has an imaginary component it will also experience absorption as it travels through the material, however here the interface only is considered, absorption shall be included later when dealing with layer thickness. Therefore due to the law of conservation of

energy, at the interface

$$R + T = 1 \tag{3.11}$$

where R is the reflectance and T the transmittance.

For normal incidence

$$R = |\Gamma|^2 = \left(\frac{E_{0r}}{E_{0i}}\right)^2, \tag{3.12}$$

$$T = \frac{n_t}{n_i}|\Im|^2 = \frac{n_t}{n_i}\left(\frac{E_{0t}}{E_{0i}}\right)^2 \tag{3.13}$$

Where subscripts t, r and i represent the transmitted, reflected and incident components respectively, Γ and $\Im$ are the reflection and transmission amplitudes and n is the media refractive index ($n = \sqrt{\varepsilon}$ where ε is the effective permittivity of the media).

For a general transmission problem at an interface there are incoming and outgoing waves on both sides. This is detailed in Figure 3.2. The transfer matrix expresses the waves on the right of an interface as a function of those on the left:

$$\begin{pmatrix} C \\ D \end{pmatrix} = \mathtt{T}^{(21)} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} T_{11}^{(21)} & T_{12}^{(21)} \\ T_{21}^{(21)} & T_{22}^{(21)} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \tag{3.14}$$

The T-matrix components are the standard reflection and transmission equations for an interface [78]. The electric and magnetic flux must be continuous across the interface. If two of the amplitudes are known, then from this the others can be deduced. Equation 3.14 can be extended to multiple interfaces by multiplication of the T-matrices. However for a structure with multiple interfaces each T-matrix must be translated to account for the displacement z_1 of the interface from the origin ($z = 0$):

$$\mathtt{T}(z_1) = \begin{pmatrix} e^{-ik_2 z_1} & 0 \\ 0 & e^{ik_2 z_1} \end{pmatrix} \mathtt{T}(0) \begin{pmatrix} e^{ik_1 z_1} & 0 \\ 0 & e^{-ik_1 z_1} \end{pmatrix} \tag{3.15}$$

The T-matrix for the multiple interface system can now be calculated:

$$\mathtt{T}^{(31)} = \mathtt{T}^{(32)}(z)\mathtt{T}^{(21)}(0) \tag{3.16}$$

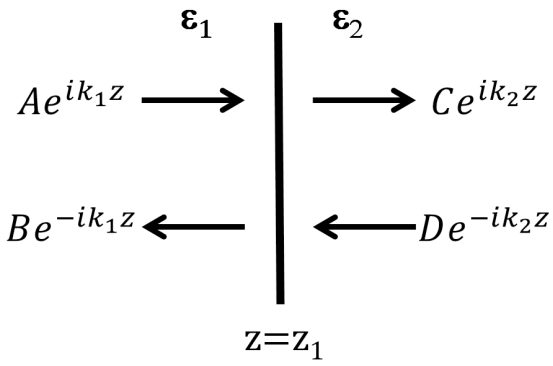


Figure 3.2: A schematic of the incoming and outgoing waves at the interface of two media with effective permittivity ε_1 and ε_2.

Γ and $\mathcal{T}$ can then be recovered from the system T-matrix by setting $A = 1$ and $D = 0$. For Equation 3.14, this equates to

$$\Gamma = -\frac{T_{21}}{T_{22}}, \qquad \mathcal{T} = \frac{T_{11}T_{22} - T_{12}T_{21}}{T_{22}}. \tag{3.17}$$

As can be seen this is a much less computationally demanding process, however the application is limited to simple 1D layer structures. It can also be easily integrated with other scripted analytical processes such as the computation of a material effective medium, as shall be discussed in more detail later in this Thesis.

3.2 Optimisation Algorithm

A brute-force trial and error method is not an efficient method for optimising a filter design with many parameters, and therefore a large search space. In such cases an optimisation algorithm, which explores the problem search space more intelligently in order to find a solution that represents the global maximum or minimum, is more suited. There are many optimisation techniques that can be applied to such design situations, these solutions fall into two main categories of local or global optimisers. Local optimisers often place constraints on the solution domain; such as the existence of a derivative (in gradient techniques). Taking advantage of such solution space characteristics enables fast convergence

to local optima. Global techniques however place fewer constraints on the solution domain and are therefore more robust but slower in optimisation. The design problem must be examined to understand which method is appropriate. The problem explored in this work is the design of a plasmonic filter to manipulate the transmission spectrum to achieve a desired passband performance. This problem has multiple geometrical variables and assessment parameters, with little prior knowledge of the search space assumed. For example no assumptions were made as to whether it has non-differentiable or non-continuous search space regions; or whether the search space presents multiple local or global optima. For such a problem a robust, global optimiser - independent of the starting conditions - is required.

Two algorithms that are particularly effective in such cases are the genetic algorithm (GA) [80, 81, 82] and simulated annealing (SA) [83, 84]. The roots of the SA algorithm are in thermodynamics; modelling the annealing process of liquid-state metals to a solid-state. The main principles of a GA are based in the theory of natural selection. For the design process in this work, the GA was selected. In comparison to SA, the GA has the ability to provide solutions when fewer parameters are known. In addition it has been noted that the performance of the SA initially has a higher dependence on the selection of the correct algorithm parameters for the specific application [85], and may require significant tuning of these parameters.

The ability of the GA to implement multiple variables, with little or no editing of the basic GA process, suggest the GA is the more adaptable of the solutions. This would be useful for other design problems in future. The GA technique is particularly effective in applications with a large number of disparate parameters; this would be applicable for implementation in the proposed design of the geometrical parameters, but also for other possible considerations in future such as material selection, or cost and manufacturing issues.

3.2.1 Genetic Algorithm

GA have been applied successfully to solve many electromagnetic design problems such as the design of frequency selective surfaces [86, 87, 88]. The algorithm

produces a population of solutions that evolve towards an optimum, while some diversity is maintained in order to ensure broad exploration of the search space. This is achieved through the use of reproduction and mutation processes, and weighting of the suitability of the solution. The GA procedure is detailed extensively elsewhere [80, 81, 82]; here only a summary of the main parts is provided. A GA proceeds through several key phases:

1. **Initiation** - An initial population of randomly-selected solutions are encoded. Each solution is referred to as a chromosome. Each chromosome is the combination of several variables (genes).

2. **Evaluation** - Each of the population's chromosomes are individually evaluated for suitability to solve the problem at hand. Their adherence to specified criteria is assessed and this is quantified in a positive value known as the solutions fitness.

3. **Crossover** - Two parent chromosomes are selected from the population. From this, child chromosomes are created, consisting of a combination of the parents chromosomes. The child chromosomes are used to form the new population.

4. **Mutation** - Bits of the new population of chromosomes are randomly mutated, with a specified probability (p_m).

The process is repeated from step two (evaluation) with each new population until a specified number of iterations of the algorithm has been completed or an adequate solution is found. A block diagram of the implementation and sequence of phases is shown in Figure 3.3.

Many variations on this simplified GA process have been presented to provide further control over the evolution of the search algorithm. For this project, a steady-state GA replacement strategy was applied: only a proportion of the population is replaced between generations, and that proportion is determined by the probability of crossover (p_c). This enables faster convergence in many applications [80]. The typical values of p_m and p_c are $p_m = 0.01 - 0.1$ and $p_c = 0.6 - 0.9$ [80, 81]. Binary is the most commonly applied encoding technique

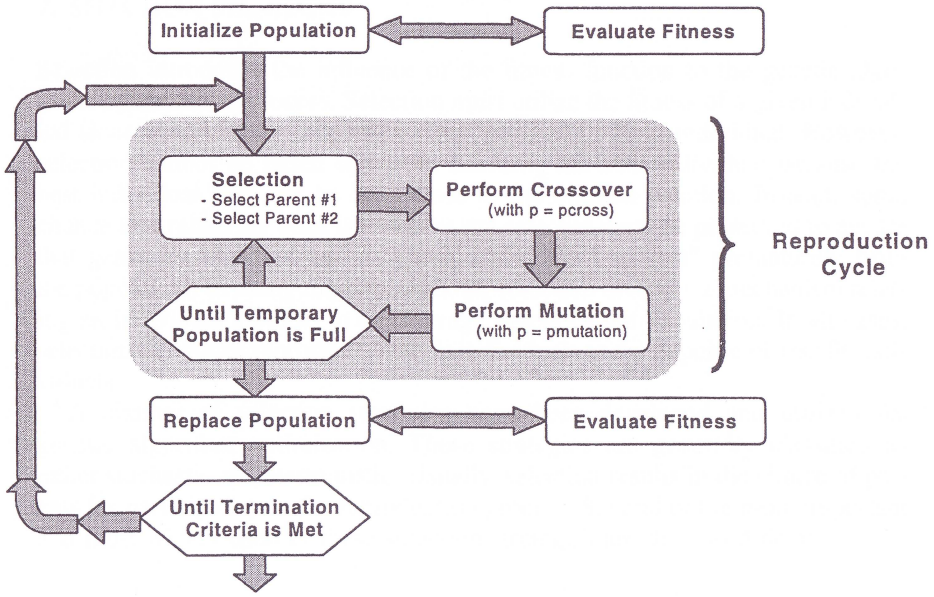


Figure 3.3: A block diagram of the genetic algorithm optimisation process [80].

because it is straightforward to discretise a variety of parameters such as direct conversion from decimal to binary of structure parameters, or the geometry of a metal layer by discretising the area into a checkerboard and using '0/1' to denote the absence/presence of the metal (as often applied in frequency selective surface design).

There are many techniques for population decimation and parent selection presented in literature [80, 81], each with varying degrees of complexity and noted benefits. For the GA in this application, a simple system of truncation population decimation is utilised; this process sorts the solutions in rank order of the fitness value and a specified number of the least fit solutions are eliminated. These spaces are filled with the child chromosomes from the reproduction process. There are also other proportionate selection techniques available including roulette wheel or tournament selection, where the solutions with the better fitness have a higher probability of selection but lower fitness solutions can still be picked (so maintaining an element of diversity). Initial results utilising the rank procedure were successful therefore the additional complexity of proportionate selection techniques was considered unnecessary.

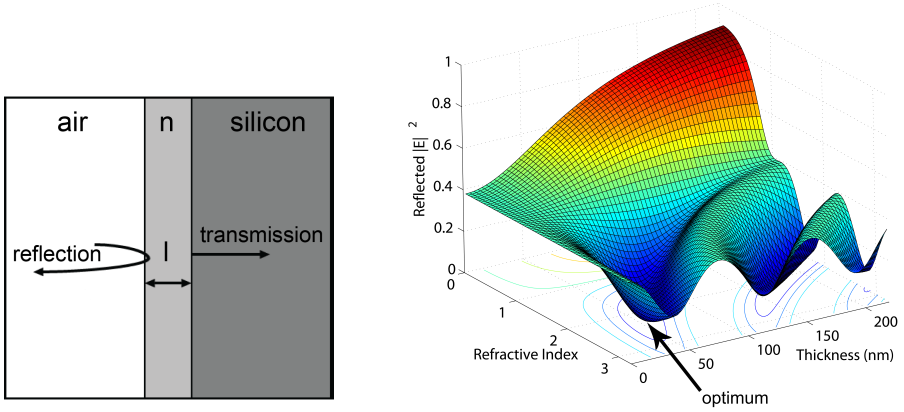
After population decimation, the eliminated solutions are replaced with child chromosomes derived from parent chromosomes. A random parent-selection process was implemented where any of the remaining solutions present, after population decimation, has a uniform possibility of selection to be a parent. Again an adaptation to this could involve a proportionate selection process. The forming of the child, from the crossing over of parent chromosomes, can be achieved via single, multiple or uniform crossover points [81], with the crossover point selected randomly. For this application a single crossover point appeared sufficient.

Many other methods have been presented to further improve and tune the performance of a GA. For example, niching [89, 90] or pareto GA [81] can investigate multiple optima, and elitism can be applied to ensure the survival of the fittest solution [80]. These variations can be explored in future, however seemed unnecessary for the level of complexity required by this design process.

3.2.2 Initial Testing

The GA search method developed was tested using a simple electromagnetic design problem, the design of an anti-reflection layer, to ensure its ability to identify an optimum solution. Additionally the process was explored to identify values of p_m and p_c that could be applied. The test structure is shown in Figure 3.4(a). This particular test problem was selected as the optimum refractive index (n) and material thickness (l), that provide the minimum reflection, can be calculated analytically [78]. A graphical representation of the search space is shown in Figure 3.4(b). The known optimal design is marked. For comparison a similar design procedure using a SA algorithm was also explored. Some literature [85, 91] has identified that SA algorithms are much slower in CPU time than GA in finding an optimum solution. This was supported by an initial study conducted for this work. For this test application the GA performed approximately 50 generations (considering this in terms of the number of FDTD simulations, as shall be applied in this work, in real time this equated to 50 times the length of an FDTD evaluation) and 950 function evaluations. The SA, although only requiring 270 function evaluations to identify the solution, in real time this equated to 270 times the length of the FDTD simulation time, thus is significantly slower in these terms.

This efficiency will be reduced if the population size of the GA is greater than the number of available processors. In such cases multiple simulation batches must be run for each generation of the algorithm. For this project, a 320 core cluster was available, so the parallel nature of the GA was not an issue.

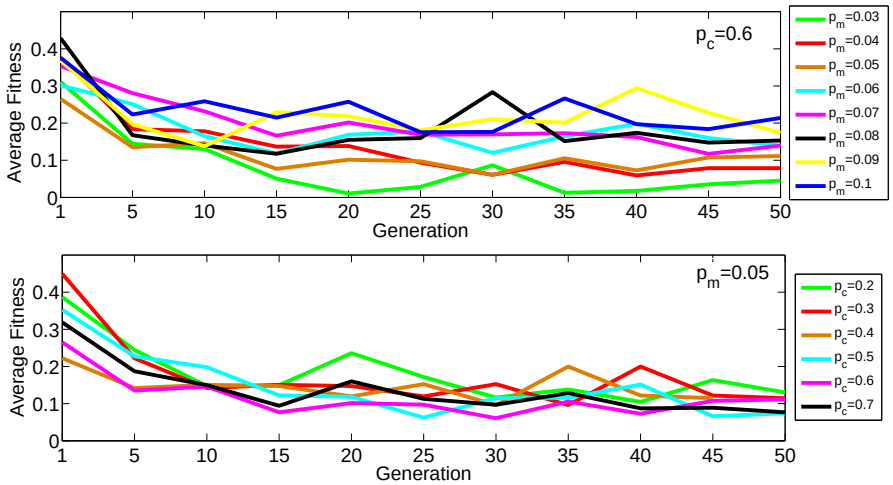


(a) (b)

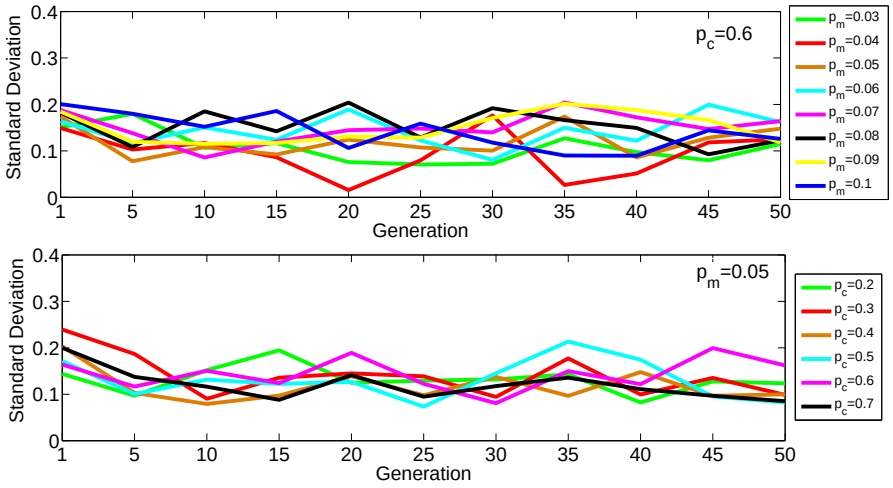
Figure 3.4: The air-material-silicon anti-reflection layer test structure with varying material thickness and refractive index (a) a schematic diagram and (b) a plot of the reflection from the air-ARL interface for the search space.

This initial trial demonstrated a GA process that ranked solutions and randomly selected parents, could be applied successfully to the optimisation of a simple electromagnetic problem. For this application, it was deemed more efficient than the SA. An exploration of p_m and p_c values was also conducted using this problem to identify suitable settings for future applications of the GA. The range $0.03 \leqslant p_m \leqslant 0.1$ and $0.2 \leqslant p_c \leqslant 0.7$ was explored and the mean and standard deviation of the fitness value of the population, at each generation of the GA procedure, is shown in Figure 3.5. It was desirable for the mean value to reduce with each increasing generation while simultaneously maintaining a high standard deviation. This would indicate a trend towards finding the optimum solution (in this case the minimum reflection) while also having a wide range of solutions present in the population. This would continue exploration of the search space, therefore increase the probability of locating the global optimum solution. From this simple examination of the performance of the optimisation procedure it can be identified that values of $p_m = 0.05$ and $p_c = 0.6$ would be

suitable for application in this GA. While a plasmonic device is more complicated than a simple anti-reflection coating; tuning the algorithm on a simple related problem is still likely to enhance its performance when applied to a problem with an unknown optimal outcome.



(a)



(b)

Figure 3.5: Plots of the results of the GA for the test structure for different values of p_c and p_m; (a) the mean fitness value (b) the standard deviation of the population at selected iterations of the algorithm.

3.3 Measurement

The measurement of the spectra of the filters was conducted using a TFProbe MSP300 spectrometer from Angstrom Sun Technologies Inc, shown in Figure 3.6. A spectrometer uses a prism or grating to focus a section of the wavelength spectrum onto a sensor. Using this technique, narrowband measurements of the filters' spectra can be conducted. This tool has a resolution of 1 nm. The tool can be used to obtain both transmission and reflection spectra. In this work, the filters were fabricated on to transparent substrates, in order that the transmission mode could be used, akin to how the filters would work on-chip. The reflection spectra may be useful for characterisation of the filters if working on non-transparent substrates such as silicon, but not used here.

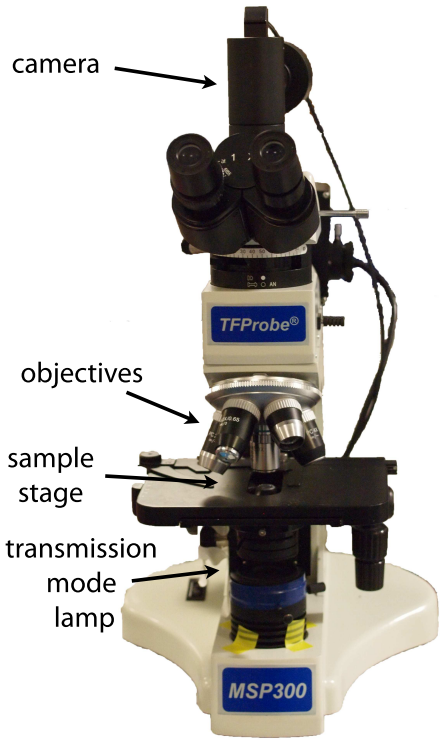


Figure 3.6: The TFProbe MSP300 spectrometer used for measurement of the filter spectral response.

The light source is a 50 W halogen bulb. This illuminates the sample structure from below. All the filter structures were fabricated directly onto a standard

microscope glass slide due to its transparency at all visible wavelengths. The incident illumination is transmitted through the structure and the collected by the spectrometer objective as shown in Figure 3.6. For the measurements conducted here an objective with 50× magnification and a spot size of approximately 10 μm was used.

The measurement process begins with a measurement of the spectra when no light is present on the sensor. The user then adjusts the illuminant level to ensure the sensor is not saturated. The transmission through the glass slide (with no structure) is measured. This provides a dark noise level and enables normalisation at the later stages to remove any contribution of the substrate and illuminant to the structure transmission spectra. When this has been conducted the filter spectra can be analysed by aligning the structure to be measured with the objective using the built-in camera. The measurement of the transmission is then conducted. This is automatically normalised to the initial calibration measurements by the spectrometer software. This results in the output spectrum being a measure of the filter transmission structure for the area of the spot size of the objective. The data is exported from the tool in a text file format.

Chapter 4

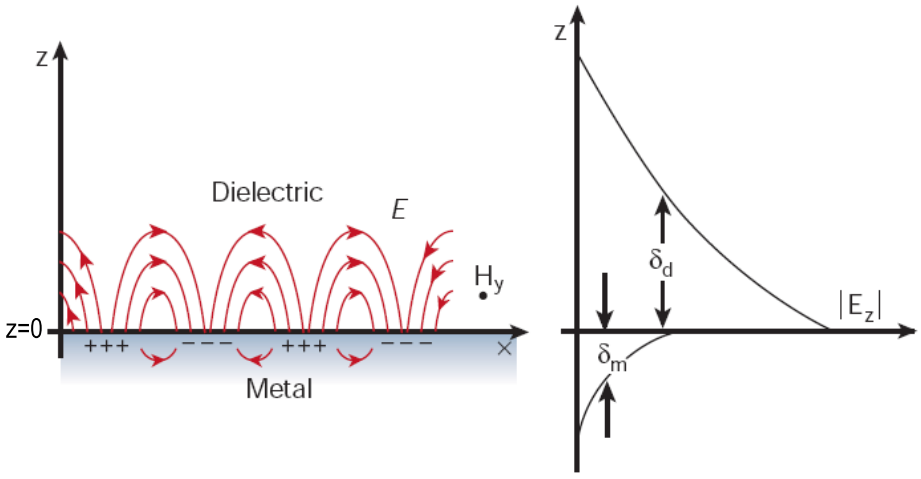
Plasmonic Filters

4.1 Introduction

Surface plasmons are collective oscillations in charge at the interface between a metal and dielectric. They can take several forms including localised oscillations on nanoparticles and propagating electron density waves. The unique capabilities of surface plasmon based structures has led to significant growth in the field of plasmonics, with applications in optical data storage [92], lenses [93, 94] and biological sensing [95, 96, 97] to name but a few. The ability to employ plasmonic metal nanostructures to guide and manipulate light confined to the interface (including manipulation of the transmission of the light incident on the structure), by the variation of the geometry of the nanostructure, has led to the development of plasmonic filters. Plasmonic filters provide an alternative filtering solution for application in CMOS image sensors. This ability to tune the filtering response is exploited in this work to produce three filters with transmission spectra matching the colour matching functions that approximate the responses of the human eye. These are referred to here as the red (600 nm), green (555 nm) and blue (445 nm) filters. Such filters could be applied in trichromatic imaging to reduce the colour reproduction error. An advanced design algorithm is applied to achieve this. In this Chapter, relevant aspects of the theory of plasmonics are summarised and the automated design procedure is detailed. Then, simulation results and the experimental measurements of the filter structures are presented.

4.2 Surface Plasmon Polaritons

A surface plasmon polariton (SPP) is an electromagnetic excitation that propagates like a wave at the interface between a dielectric and a metal. They occur as a consequence of the coupling of electromagnetic fields to the free electrons of the metal, causing them to oscillate in resonance as shown in Figure 4.1(a). The SPP's electromagnetic field is confined to the interface and the amplitude decays exponentially in the perpendicular direction (evanescently confined) as demonstrated schematically in Figure 4.1(b). Here δ_d and δ_m are the penetration depths into the dielectric and metal respectively. This confinement at the interface results in the SPP being extremely sensitive to changes in the surface conditions. It is this property that is exploited in some of the applications of plasmonics, such as biological sensing.



(a) (b)

Figure 4.1: A surface plasmon propagating (in the x-direction) at the interface between a metal and dielectric; (a) a schematic of the associated electromagnetic field at the interface between a metal and dielectric and (b) the evanescent field perpendicular to the interface (evanescent decay in the z-direction)[98].

A planar interface between a metal and a dielectric is the simplest geometry that can sustain a SPP, where the dielectric has a real positive dielectric constant ε_d, and the metal has a dispersive dielectric constant defined $\varepsilon_m(\omega)$. SPP only exist for TM polarisation as detailed in [95]. From Maxwell's equations, the TM

solutions in the dielectric (i.e. $z < 0$ if the interface is $z = 0$ and homogeneous in the y direction) can be defined as:

$$H_y(z) = A_d e^{ik_{x,d}x} e^{k_{z,d}z} \tag{4.1}$$

$$E_x(z) = -iA_d \frac{k_{z,d}}{\omega\varepsilon_0\varepsilon_d} e^{ik_{x,d}x} e^{k_{z,d}z} \tag{4.2}$$

$$E_z(z) = -A_d \frac{k_{x,d}}{\omega\varepsilon_0\varepsilon_d} e^{ik_{x,d}x} e^{k_{z,d}z}, \tag{4.3}$$

and in the metal ($z > 0$):

$$H_y(z) = A_m e^{ik_{x,m}x} e^{-k_{z,m}z} \tag{4.4}$$

$$E_x(z) = iA_m \frac{k_{z,m}}{\omega\varepsilon_0\varepsilon_m(\omega)} e^{ik_{x,m}x} e^{-k_{z,m}z} \tag{4.5}$$

$$E_z(z) = -A_m \frac{k_{x,m}}{\omega\varepsilon_0\varepsilon_m(\omega)} e^{ik_{x,m}x} e^{-k_{z,m}z}. \tag{4.6}$$

Where $k_{z,d}$ and $k_{z,m}$ are the components of the wave vector in the direction perpendicular to the interface, in the dielectric and metal respectively. The component of the wave vector parallel to the interface is conserved thus the subscript defining the medium is unnecessary ($k_x = k_{x,m} = k_{x,d}$). H_y, E_x and D_z are continuous at the interface therefore $A_d = A_m$. From this we can define

$$\frac{k_{z,d}}{k_{z,m}} = -\frac{\varepsilon_d}{\varepsilon_m(\omega)}. \tag{4.7}$$

In addition H_y must fulfil the wave equation [95] therefore from this, using Equations 4.1 and Equation 4.4 respectively, the following relations can be derived:

$$\begin{aligned} k_{z,m}^2 &= k_x^2 - k_0^2\varepsilon_m(\omega) \\ k_{z,d}^2 &= k_x^2 - k_0^2\varepsilon_d. \end{aligned} \tag{4.8}$$

The combination of Equations 4.7 and 4.8 enables the dispersion relation of the

SPP propagating at the interface to be defined:

$$k_{\mathrm{SPP}} = k_0 \sqrt{\frac{\varepsilon_d \varepsilon_m(\omega)}{\varepsilon_d + \varepsilon_m(\omega)}} \tag{4.9}$$

The real parts of $k_{z,m}$ and $k_{z,d}$ must be positive to describe a wave confined to the interface. Therefore we can see from Equations 4.7 that if the real part of ε_d is greater than 0 then the real part of $\varepsilon_m(\omega)$ is less than 0, confirming the surface waves only exist between a conductor and dielectric. The requirement that $k_{z,m}$ must be real and positive results in the SPP dispersion curve lying to the right of the dielectric medium light dispersion as shown in Figure 4.2. As a consequence of this the SPP can not be excited by light incident on, or radiate light into, the dielectric medium at a planar interface. This mismatch in the wave vector must be overcome to enable incoming light to couple to a SPP. This can be achieved through the illumination at an angle using a prism or alternatively, as shall be used in this application, by diffraction effects. This can be achieved by the use of topological features such as subwavelength holes or periodic gratings [95]. Components of the diffracted light whose wave vectors match that of the SPP will be coupled to SPP.

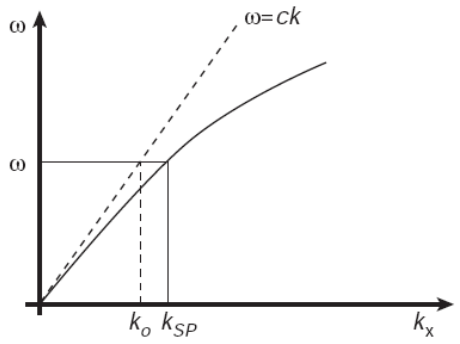


Figure 4.2: The dispersion curve of the SPP at an interface (solid) and the dispersion of light in a dielectric medium (dashed) [98]. For a given ω_0, $k_{\mathrm{SP}} > k_0$

For a grating, the momentum matching condition occurs when

$$\begin{aligned} k_{\mathrm{SPP}} &= k_{||} + iG_x \\ &= k_0 \sin\theta + i\frac{2\pi}{a} \end{aligned} \tag{4.10}$$

where $k_{||}$ is the in-plane wavevector of the incident light, G_x is the reciprocal vector of the grating, a is the grating period and i an integer. The grating can also be applied to reverse the process, where the SPP is coupled to light, therefore re-radiating. Using this, the wavelength of SPP excitation can be varied by altering the grating period.

If the metal is a layer sandwiched between two dielectrics then SPP can occur at both metal-dielectric interfaces. This can be exploited; if the thickness of the metal layer is of the order of a skin depth and dielectric upper and lower cladding are equal, then coupling of the SPP on the adjacent interfaces can take place. This enables phase matching and enhancement of the tunnelling through the aperture [95].

4.2.1 Hole Arrays

The structure that has shown particular promise for plasmonic filtering is that of an array of subwavelength holes in a metallic layer. The extraordinary transmission properties of such a structure were first demonstrated by Ebbesen *et al* [99]. Extraordinary transmission is where the percentage of light transmitted is greater than the percentage of hole area to total area [95]. Although note that some rationalisation of the enhancement factor has since occurred because early work assumed too little light got through an isolated hole. The formation of SPP on such structures also enables filtering of the transmitted wavelengths, therefore forming a compact visible-wavelength filter.

For subwavelength structures ($\lambda > 2d$, where here d is the hole diameter), no propagating modes can exist in the hole. Several explanations of the role of SPP in the transmission have been presented. Krishnan *et al* [100] explain the transmission of the incident light as a consequence of formation of the SPP at resonance, resulting in the interfaces acting like "supermirrors" for the evanescent

waves. This results in constructive interference. This is likened to a Fabry-Pérot cavity, except in this instance for evanescent waves, where the holes form subwavelength cavities. Maier *et al* explain that the SPP lead to an enhanced field on top of the aperture, which is coupled through the aperture and re-radiates on the other side.

The model adopted by Rivera *et al* [101] for a single slit involves multiple coupling processes. In this model they explain that the SPP formed on the upper interface reconverts to a propagating wave at the hole, therefore interfering with the field directly transmitted through the hole. They also claim localised surface plasmons [95] are formed on the rim of the hole, and FP resonances with the apertures, contribute to the modulation of the transmission. As can be identified many theories are presented for the exact role of the SPP in the transmission through the structure. These different theories present various relations between the surface plasmon resonance and the transmission spectra of the structure.

The resonant wavelength of the surface plasmon can be derived analytically. The coupling of SPP is achieved by the diffraction effects of the light incident on the periodic structure. As with a 2D grating the 3D hole structures result in components of the diffracted light having wavevectors matching that of the SPP, thus overcoming the momentum mismatch. Similar to Equation 4.10, the appropriate surface plasmon wave vector is achieved if the grating conditions match that of

$$k_{\mathrm{SPP}} = \sqrt{\left(k_{||}\cos\phi + i\frac{2\pi}{a_x} + j\frac{2\pi}{a_x}\right)^2 + \left(k_{||}\sin\phi - i\frac{2\pi}{a_y} + j\frac{2\pi}{a_y}\right)^2} \tag{4.11}$$

where a_x and a_y are the period in the x and y direction, i and j are integers and ϕ is the angle with the x-axis [102]. Combining Equation 4.9 and 4.11, the wavelength at which the SPP resonances are predicted to occur are

$$\lambda_{\mathrm{SPP}} = \frac{2\pi\sqrt{\frac{\varepsilon_m\varepsilon_d}{\varepsilon_m+\varepsilon_d}}}{\sqrt{(k_{||}\cos\phi + i\frac{2\pi}{a_x} + j\frac{2\pi}{a_x})^2 + (k_{||}\sin\phi - i\frac{2\pi}{a_y} + j\frac{2\pi}{a_y})^2}} \tag{4.12}$$

Equation 4.12 has in the past been used to predict the transmission maxima [95, 103]; the filter transmission peaks being attributed to the SPP resonant

excitation. Equation 4.12 is in fact a simplification of a complex phenomenon. Equation 4.12 is based on a planar, solid interface and does not account for the influence of the holes on the permittivity or SPP resonance. Or indeed, that the hole could couple SPP modes between the surfaces. This particular feature is cited by some [104] as resulting in a shift to shorter wavelengths of the peak from that predicted using Equation 4.12, as commonly observed in simulation and indeed experimental results. More recently it has been proposed that the interference between the direct transmission and the surface mode results in the red shift [105]. This equation therefore would in fact relate to the minima that results from the destructive interference [105, 106]. In particular [106] notes the inadequacy of the application of Equation 4.12 to predict maxima as there is no incident light on the lower interface (which will determine the transmission). This therefore only predicts the radiation if the lower interface SPP were identical to those on the top interface. In addition they note similar features show up in aperiodic arrays with no well-defined grating momentum, hence they claim that short range order is responsible for the formation of transmission spectra . The suggestion that it is the transmission minima that is predicted by the SPP wavelength (Equation 4.12) is consistent with the observations in this work (Figure 4.10).

Further features of the transmission spectrum, such as the peak width and amplitude, cannot be predicted by Equation 4.12. The attenuation of the transmission is influenced by the material properties, the hole diameter, the SPP field intensity at the interface and the SPP lifetime [105]. These are difficult to predict analytically. Similarly the width of the peak is dominated by the radiative scattering of the SPP [105]. Nonetheless, trends can be identified. For example, in thinner metal layers, where coupling can occur between the SPPs on the upper and lower interface, the transmission peak broadens with reducing metal thickness [107, 108]. Similarly, if the upper dielectric layer's thickness increases, or the hole diameter is increased, the transmission peak increases and broadens. From this it can be identified that the filter transmission can be manipulated by variation of the SPP wavelength, propagation length and scattering. This is determined by the geometrical features and the dispersive behaviour of the materials used. As there are many variables in the design that influence the specific filtering performance, an advanced optimisation procedure is required. To

attempt to design such filters to replicate the CMF, as shown in Figure 2.2, a combined search algorithm and FDTD simulation procedure was required. Such a methodology was described in Section 3.2.

4.2.2 Structure

A set of *rgb* filters based on a plasmonic structure using a triangular period array of holes, as shown schematically in Figure 4.3, has previously been demonstrated experimentally by Chen *et al* [108]. This structure was used as the basis for the development of the CMF filters.

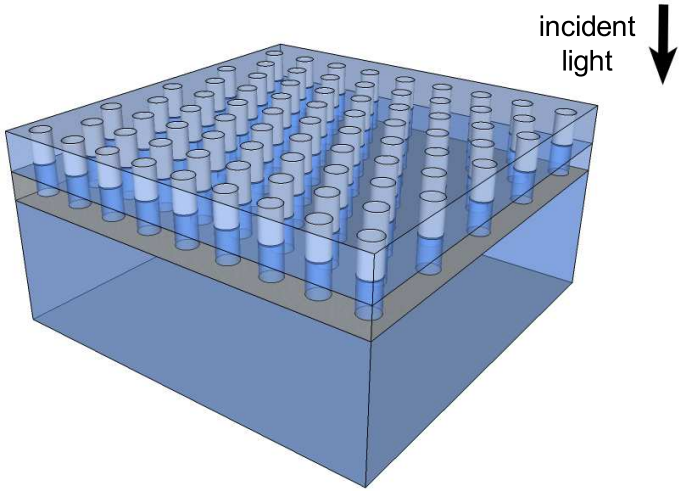


Figure 4.3: Diagram of a plasmonic filter hole filter array with a triangular periodicity.

The metal in a plasmonic filter must ideally have a large real part of the dielectric function in addition to a minimal or zero absorption component. Gold and silver are commonly used because they have excellent properties for supporting SPP in the visible range. Gold and silver however are not compatible with the CMOS process and therefore could not be applied in this work. Instead, aluminium (Al) was used because it provided suitable material properties for SPP resonance, in addition to being CMOS compatible and has good adherence to SiO_2 and Si. The disadvantage of Al is slightly increased absorption compared to gold or silver.

The filters are fabricated onto microscope slides for testing, the design process therefore assumes a semi-infinite SiO_2 substrate. On top of this is an Al layer with a periodic hole pattern etched through the thickness of the Al. This is covered by an additional SiO_2 layer; this served to match the dielectric constant at the upper and lower interface of the structure. The resulting in-filling of the Al layer holes, and therefore formation of holes in the top layer as would be expected after deposition of the upper layer, is also included in the design. This is demonstrated schematically by the cross-section of the structure in Figure 4.4(a).

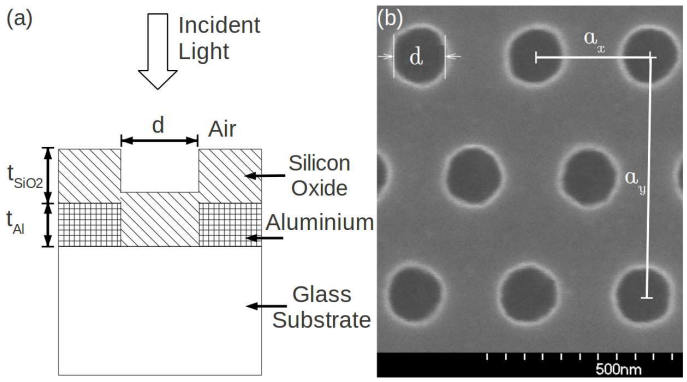


Figure 4.4: Plasmonic filter structure (a) schematic of a single period of the filter structure. The holes etched into the aluminium, deposited on a glass substrate, are filled by the SiO_2 top layer. (b) A Scanning Electron Micrograph (SEM) image of an array of holes etched into the aluminium identifying the array parameters.

Plasmonic filters using these materials had previously been optimised and successfully demonstrated by our research group [108]; as a consequence of this prior work the fabrication procedures had also been optimised. For this application the same filter design was applied. To enable tuning of the transmission performance the design procedure could vary the upper dielectric and aluminium layer thickness (t_{SiO_2} and t_{Al}), and the hole diameter (d). The ability to tune the lattice period (a_x and a_y) would enable designs with a rhombic or triangular lattice.

4.3 Optimisation Procedure

An advanced optimisation procedure was created to design the filters by searching the large parameter space of the structure shown in Figure 4.4, and identifying an optimal solution to meet the desired specified criteria of a transmission spectra matching that of the CMF. The GA, as previously detailed in Section 3.2, was integrated with a FDTD simulation process, to achieve a fully automated design optimisation procedure. This process repeats a cycle of selecting the parameters of a defined structure, simulating the optical performance and analysing the results, with the aim to evolve the design towards an optimum.

A block diagram outlining the different stages of the optimisation procedure is shown Figure 4.5. The process begins with the user initialising the design procedure; identifying the desired number of GA iterations, algorithm variables and the number of available processor cores available for simulation as arguments. The design procedure is then fully automated with the algorithm optimising the design with no further input required by the user. As the design parameters, simulation file and fitness analysis are highly dependent on the structure and the desired criteria, these are separate process functions that are edited to the user requirement. The remainder of the optimisation process is generic and can be implemented with little or no editing.

The algorithm begins by generating a population of random binary sequences which are translated into a design geometry. These parameters are encoded into input files that represent the desired structure, in the format required by the simulation software. These files are then automatically submitted to the FDTD simulator for computation. On completion, an evaluation of the transmission performance of each solution is returned. This data is transferred into the optimisation algorithm. In addition this value, and the associated design parameters, are recorded in a datafile. This records the algorithm development and also enables identification of the final solution on completion. If the desired number of iterations have not been completed, the process will repeat.

The optimisation procedure is executed in Octave (an open source numerical computing environment). The efficiency of this design process was improved by incorporating a batching and scripting routine that enables the computation of

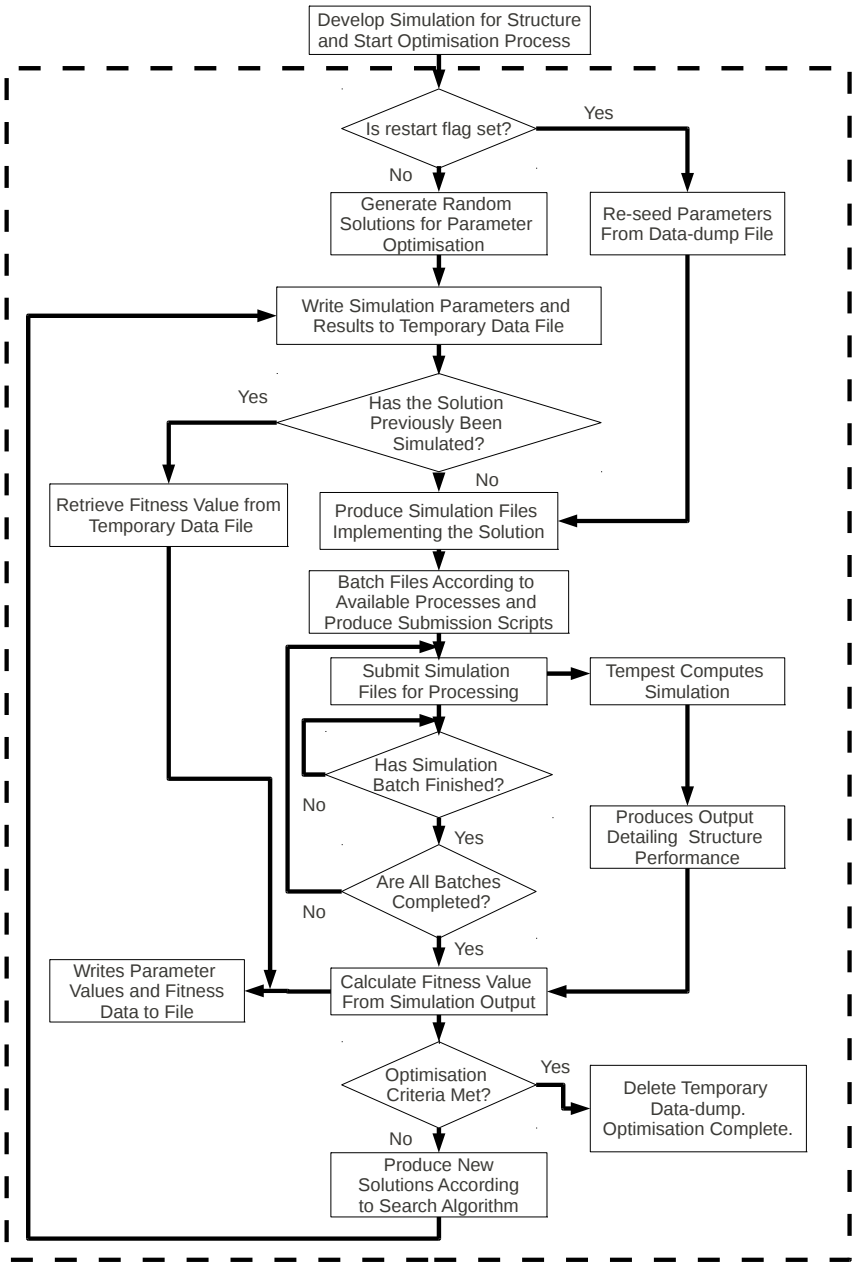


Figure 4.5: A block diagram of the optimisation process. The procedures within the dashed area are fully automated and require no user input after initialisation.

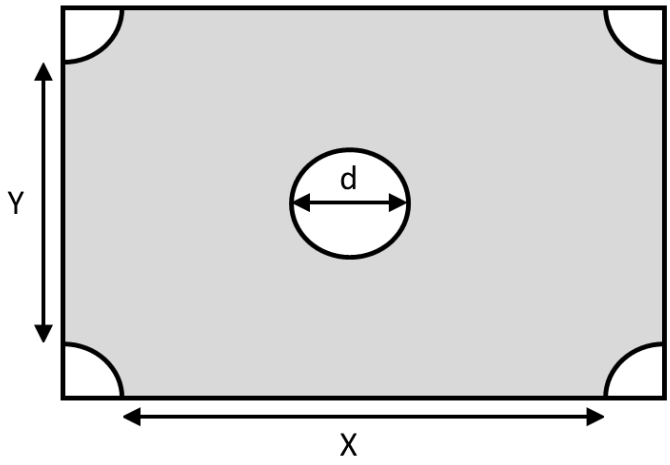
multiple simultaneous simulations on a cluster. Further to this, an adaptation was made to enable recording of the computed simulation parameters, and their associated performance. Before any batching of the files for the simulation, the

process would identify if the desired structure had previously been submitted. It would then remove this from the batching process and substitute the appropriate stored result into the data output, thus removing any repetition of simulations. Additional features such as a data-dump and reseeding process for the GA were also introduced to protect against loss of data due to unforeseen circumstances such as power cuts or machine failure.

4.3.1 Design Parameters for the Optimisation Procedure

For this application three runs of the optimisation procedure were required; each one designing a filter to match one of the three CMF. The three designs are required to have identical metal thickness (t_{Al}) and oxide coating thickness ($t_{\mathrm{SiO_2}}$) (see Figure 4.4(a)) to enable fabrication in a single lithography cycle. For this reason the first implementation had the hole diameter (d), the space between holes in both x and y directions (X and Y) and the layer thickness ($t_{\mathrm{SiO_2}}$ and t_{Al}) as search variables. Once the material thicknesses were obtained from the design of the first filter, they were then introduced as constraints in the design of the subsequent filters. The structure dimensions d, X and Y were kept as design variables. As the $\bar{x}(\lambda)$ function had two peaks, this was identified as the most difficult spectra to reproduce. This therefore was implemented into the optimisation process first.

A period of the metal layer of the filter, as viewed from above, is shown in Figure 4.6. This demonstrates how three of the design parameters (X, Y and d) relate to the chromosome produced by the GA. The number of bits, the dimension range explored and the resolution for each of the design parameters is detailed in Table 4.1. For example X, Y and $t_{\mathrm{SiO_2}}$ were set as six bit chromosomes; an increase of 1 in the binary digit corresponded to an increase in the dimension by one 10 nm cell in the simulation. This gave a search range of $10 - 640$ nm. Only four bits were used for the hole radius, therefore $d = 20 - 320$ nm, such a range would help to ensure uniform etch profiles. The algorithm parameters used were those as identified from the initial test case (Section 3.2.2), $p_m = 0.05$ and $p_c = 0.6$. The process was run for 100 iterations of the algorithm, this provided a useful rate of change of the population, so no further changes were made to p_m



Parameter	X_1 X_2 X_3 X_4 X_5 X_6	Y_1 Y_2 Y_3 Y_4 Y_5 Y_6	d_1 d_2 d_3 d_4
Chromosome	1 0 0 1 0 1	0 1 1 0 1 0	0 0 1 1

Figure 4.6: A diagram of a period of the metal layer of the filter, as viewed from above. This demonstrates how the geometrical parameters X, Y and d, are related to the chromosome as used in the GA.

Table 4.1: The limitations of the design variables used in the design procedure.

Parameter	**Bits**	**Range** (nm)	**Resolution** (nm)
X	6	$10 - 640$	10
Y	6	$10 - 640$	10
d	4	$20 - 320$	20
$t_{\mathrm{SiO_2}}$	6	$10 - 640$	10
t_{Al}	5	$10 - 320$	10

and p_c.

For the derivation of the fitness of each solution the transmission spectra, as recovered from the simulation output, was compared to the appropriate CMF function. The CMF peak was normalised to the peak value of the transmission to ensure the solution was not falsely weighted due to a lower transmission. The least-mean-square (LMS) error at each wavelength of the filter transmission spectra and the CMF was calculated to form a measure of the variation from the desired spectra, or the fitness of the solution. This value was used in the

algorithm to rank the suitability of the selected parameter in relation to the desired outcome. This assessment of fitness however was not sufficient to ensure the desired spectra. Using this measure, filters with minimal transmission (for example in one test run, $\sim 1\%$) would be identified as the best solution because the low amplitude would result in a small error. To avoid this, and to ensure a reasonable level of transmission, the minimum amplitude of the CMF was set to 25%. Solutions below this level of transmission therefore would provide a higher (poorer) fitness value. For example, a transmission peak with a maximum amplitude of 35% would be subtracted from a CMF with an maximum peak of 35%, however a solution with a maximum transmission of 20% would be subtracted from a CMF with peak value of 25%. This weighted the solutions to promote solutions with amplitudes greater than this.

4.3.2 Simulation Parameters

The assessment of the chromosome fitness value in this application is achieved by the translation of the variables into structure dimensions within a FDTD simulation file. Running this file and extracting the output would enable assessment of the transmission performance. As discussed in Section 3.1.1 there were two software options available for simulation. For this application TEMPEST was selected as this did not require individual engine licences, and therefore could support parallel simulations, as required by the GA. The FDTD algorithm in the TEMPEST code was adapted to model dispersive materials (as detailed in Appendix A); this modified version of TEMPEST was applied in the simulation of the plasmonic filters.

Input files were generated from the design parameters selected by the GA. Each simulation modelled a single 3D unit cell of an infinitely periodic structure as shown in Figure 4.7. This was bounded above and below by PML, and periodic boundaries at the sides. The illumination (source) was at the top of the simulation and propagated in the $-z$ direction. As the light from a typical scene is unpolarised, this was achieved in the simulations by using an average of two orthogonal TEM modulated Gaussian pulse sources (one parallel to the x-axis and the other to the y-axis).

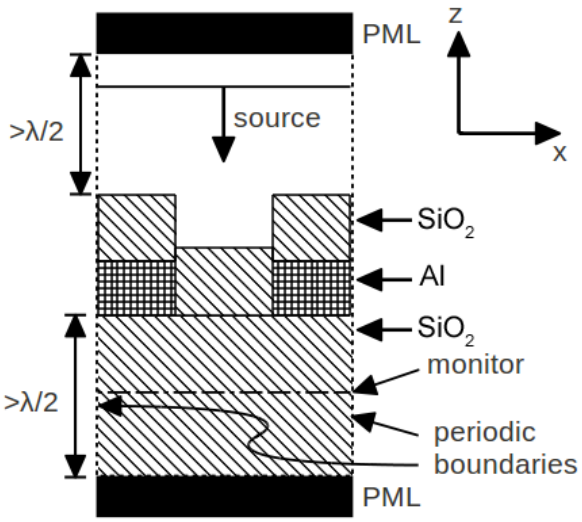


Figure 4.7: A cross-section of the FDTD domain showing a single unit cell of the plasmonic structure.

The size of the domain was altered to suit the values of a_x and a_y selected by the algorithm. The size of the simulation domain in the z-direction was altered to ensure the PML layers were at least a distance of greater than $\lambda/2$ from the material interfaces as the behaviour in the presence of evanescent fields is unreliable. The dispersive properties of the Al in the visible spectra were defined by the Drude + 2 Critical Points model (Appendix A). The refractive index of the substrate and top layer was assumed constant ($n = 1.462$) with negligible loss. The cell size was $\Delta x, y, z = 10\,\text{nm}$. A monitor was positioned in the glass substrate to record the transmission spectrum.

4.4 Design Process Results

The optimum parameters for the plasmonic filter structure, as identified by the design process, and the LMS error are presented in Table 4.2. The transmission spectra from the FDTD simulation of the selected designs are shown in Figure 4.8. These are plotted alongside the CMF normalised spectral response for comparison. Most notably the optimisation procedure was successful in the design of a filter with two transmission peaks $(440, 600\,\text{nm})$ for the red filter as desired.

Table 4.2: Filter parameters (see Figure 4.4) as identified by the design procedure. $t_{SiO_2} = 220\,\text{nm}$ and $t_{Al} = 70\,\text{nm}$ for all filters.

Filter	**d (nm)**	**a_x (nm)**	**a_y (nm)**	**LMS error**
Red	160	500	420	2.47×10^{-3}
Green	140	300	620	0.91×10^{-3}
Blue	100	380	260	1.33×10^{-3}

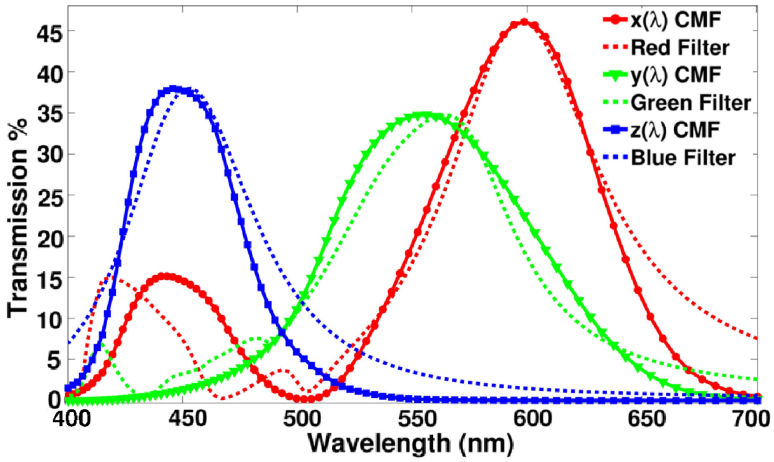


Figure 4.8: Plot of the spectral response of the filters identified as the optimum match to the CMF by the GA and FDTD optimisation process. The CMF amplitudes are scaled to match that of the simulation results. Such scaling can be adjusted trivially in a sensor by weighting the output of the RGB channels. The shape of the response is what primarily matters in determining colour error.

The simulation of the filters predicted transmission peaks of 46, 35 and 38% for the red, green and blue filters respectively. The full-width-half-maximum (FWHM) of the main peaks was 81, 75 and 61 nm. This can be compared to the FWHM of the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ CMF of 79, 100 and 55 nm. This is a difference of only 3, 25 and 6 nm in the FWHM of the filters compared to that desired. The central wavelength of the designed filters were 599, 564 and 453 nm. The red filter central wavelength matched exactly with that of the corresponding CMF. The green and blue filters had a very small red shift of only 8 and 7 nm respectively compared to that of the corresponding CMF (with central wavelengths of 556 and 446 nm).

4.5 Fabrication

A schematic diagram of the steps involved in the fabrication of the filters is demonstrated in Figure 4.9. All the filters were fabricated simultaneously on the same sample in a single lithography process. The fabrication process begins with the cleaning of the glass slide, used as the substrate, using a standard cleaning process to ensure it is clear of any contamination. The next step is the deposition of the aluminium layer using an electron beam evaporator (Plassys MEB 400S). Next a layer of ZEP520A resist was spin coated onto the sample and baked in an oven for 180 °C for 40 mins. The various hole array patterns comprising the filters were transferred to the sample using a Vistec VB6 UHR EWF electron beam lithography tool. The resist was developed using o-xylene. An O_2 ashing process was performed using a barrel asher to remove any remaining resist within the holes. The pattern is then formed in the aluminium by a dry etch procedure using a Plasmalab System 100 reactive ion etch tool. A second O_2 ash process is performed to remove the resist. The hole array of the green filter at this stage of the fabrication process is pictured in Figure 4.4(b), with $d = 139$ nm. The final step is the deposition of the upper layer of SiO_2 using an Oxford Instruments PECVD 80+. Further details of the process and recipes involved in the fabrication are provided in Table 4.3.

The electron beam (EB) lithography tool available at the University of Glasgow has a spot size of 4 nm. Although this prototyping technology is not currently available in a typical production line, such structures can be transferable to EUVL or similar tools when such tools are later integrated into the CMOS production line. Some of the nanopatterned metallic structures (with a minimum feature size of 75 nm) as presented in this work have already been fabricated using nanotransfer printing [109], therefore they could be readily incorporated immediately into the BEOL processing with the addition of such facilities.

4.6 Measurement and Analysis

The transmission of the fabricated filters was measured using the spectrometer described in Section 3.3. The measured spectra are shown in Figure 4.10(a), and

Table 4.3: Details of the processes involved in the fabrication of the plasmonic filters.

Process	Details
Resist Deposition	ZEP:Anisole=2:1, Spin at 5000 rpm for 1 min
Development	23 °C for 35 min
O_2 Ash (1)	40 W for 30 sec
Al Dry Etch	$SiCl_4$ = 18 sccm, 9 mT, 250 W (etch rate ~ 25 nm/min)
O_2 Ash (2)	200 W for 5 mins
SiO_2 Deposition	SiH_4 = 7 sccm, N_20 = 197 sccm, N_2 = 85 sccm, 1 T, 15 W, 300 °C (etch rate ~ 69 nm/min)

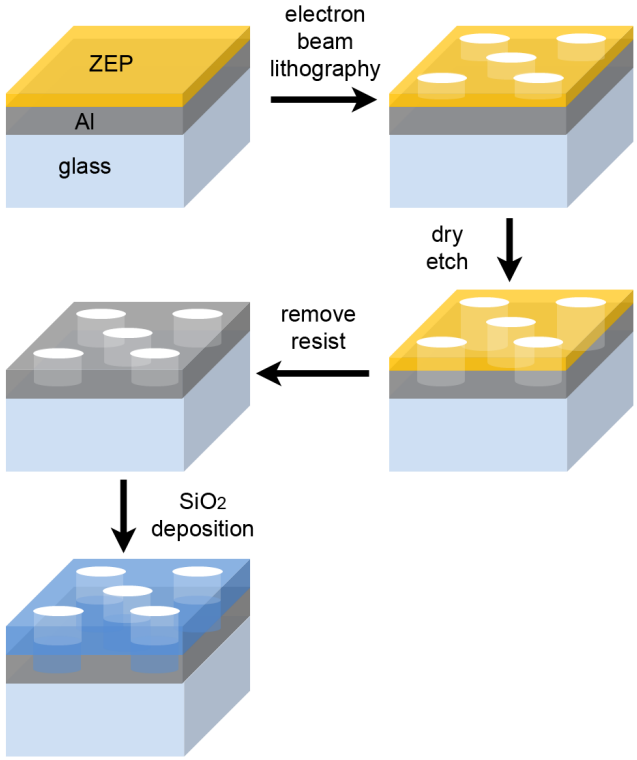
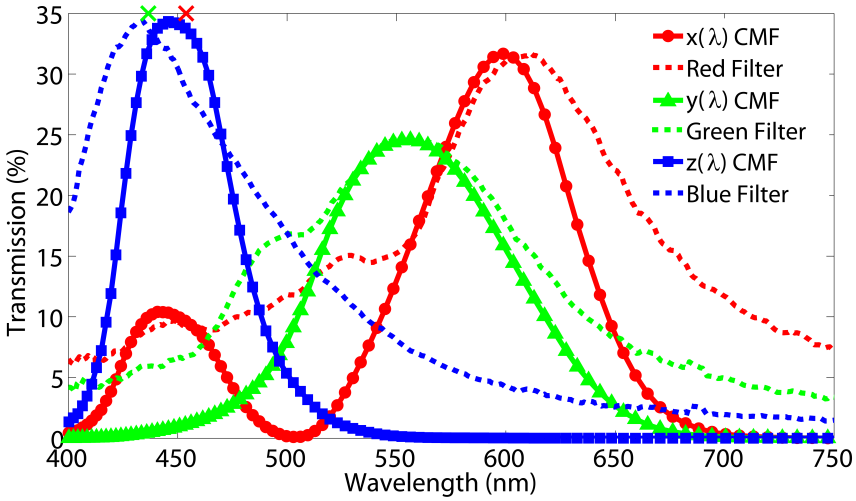


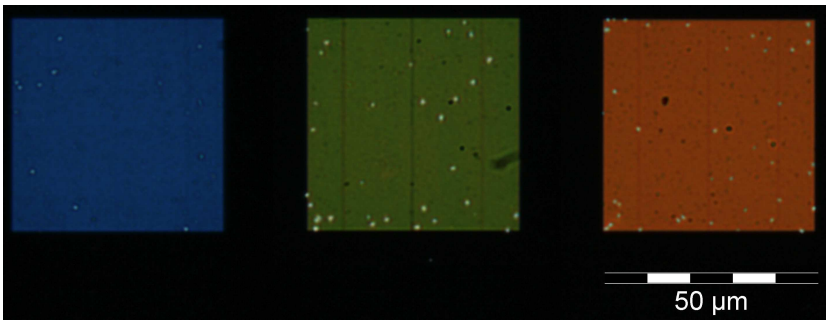
Figure 4.9: Diagram of the process flow for the fabrication of the plasmonic filters.

a white-light microscope image of the filters is shown in Figure 4.10(b). It can be identified that all the filters produced a bandpass response in the wavelength band designed. The measured transmission was 32, 25 and 35% for the red, blue and green filters consecutively, slightly reduced compared to the simulated values.

The FWHM were 117, 150 and 100 nm, or $40-75$ nm larger than predicted by simulation. For the red filter the simulation indicated that the secondary peak (at λ= 440 nm), similar to the $\bar{x}(\lambda)$ CMF, should be realisable. In this structure however, as a consequence of the broadening, this peak was not separately resolved in the experimental results. The LMS error between the measured filters and the CMF was calculated as 6.374×10^{-3}, 2.338×10^{-3} and 11.095×10^{-3}. The significant increase in the LMS of the blue filter transmission, from that achieved with the simulated filter, is attributed to the slight blue shift in the resonant wavelength.



(a)



(b)

Figure 4.10: The transmission of the fabricated filters; (a) a plot of the measured transmission spectra compared to the CMF and that predicted analytically ($\times$), and (b) a compilation of the white light microscope images of the filters.

Also plotted (as a cross) in Figure 4.10(a) is the λ_{SPP} predicted analytically using Equation 4.12 for the red and green filter dimensions ($i = j = 1$). The λ_{SPP} values calculated for the other orders, and all orders of the blue filter, are outside of the visible range. It is difficult to apply Equation 4.12 as λ_{SPP} and $\varepsilon_m(\lambda)$ are interdependent. Nevertheless from the comparison here it can be identified that Equation 4.12 did not predict the transmission maxima, instead the predicted wavelength is in the approximate region of the transmission minima. This demonstrates that the plasmonic filter behaviour cannot be predicted analytically, except for the location of transmission minima.

The variation in main peak amplitude and width could be a consequence of minor variations in the hole shape and non-zero line edge roughness. As can be seen in the SEM image of the blue filters shown in Figure 4.11, the fabrication process was successful in the formation of the hole array over a significant number of array periods with no visible defects. On closer inspection, as can be seen in the insert of Figure 4.11, the holes are not perfectly rounded and this will result in variations in the diameter and affect the performance. A measurement of two holes in the insert image identified their diameters as 100 and 107 nm. Additional simulations were conducted to explore the effect that variation in the hole size would have on the transmission spectra of the filters. The structures outlined in Table 4.2 were simulated with 50% of the holes in the array reduced in diameter by 20 nm. For example, for the blue filter 50% of the holes in the structure had $d = 100$ nm and the other 50% had $d = 80$ nm. The resulting spectra are shown in Figure 4.12. The blue and green filters are the most susceptible to the fabrication tolerances due to their smaller dimensions. The blue filter demonstrated a 13% drop in transmission, similar to the transmission reduction between the simulated and measured spectra. The FWHM was reduced to 44 nm. The transmission maximum for the blue filter was blue-shifted 10 nm to 435 nm. This blue shift is similar to the shift observed between the simulation and measured spectra. This variation of the hole diameter also resulted in shoulders on the resonances, particularly in the green, similar to those observed in the measured filter spectra.

An increase in the hole size would result in an increase in transmission and broadening of the peak. This can be observed in the set of measured filters shown in Figure 4.13. The hole diameter was varied here by maintaining the

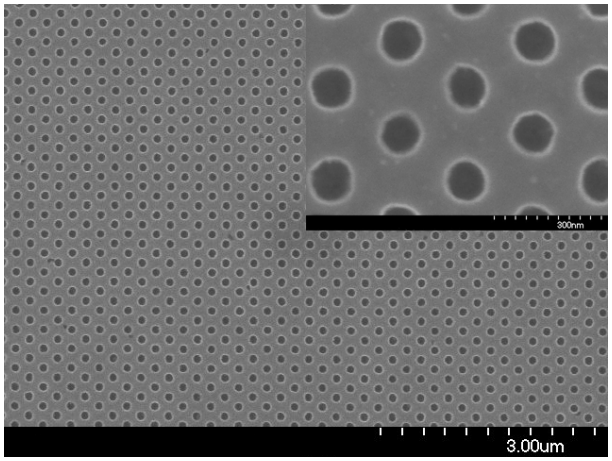


Figure 4.11: A scanning electron micrograph of the blue filter. Measurement of two holes from the insert image gave $d = 100$ and 107 nm.

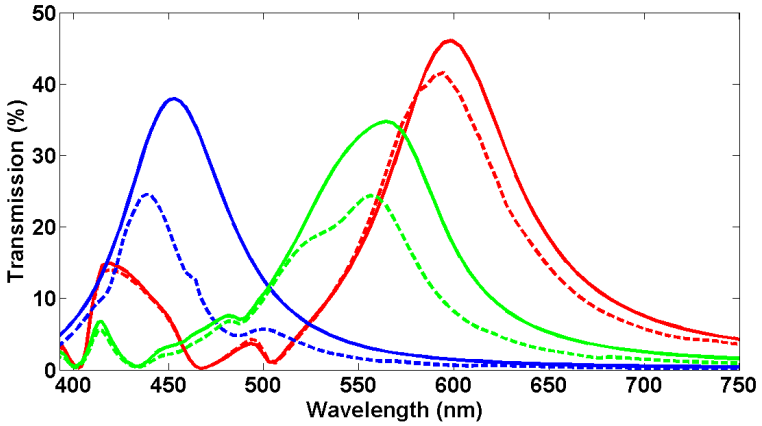
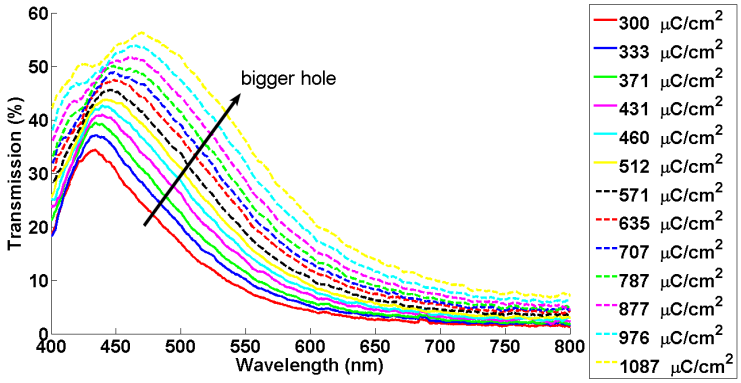


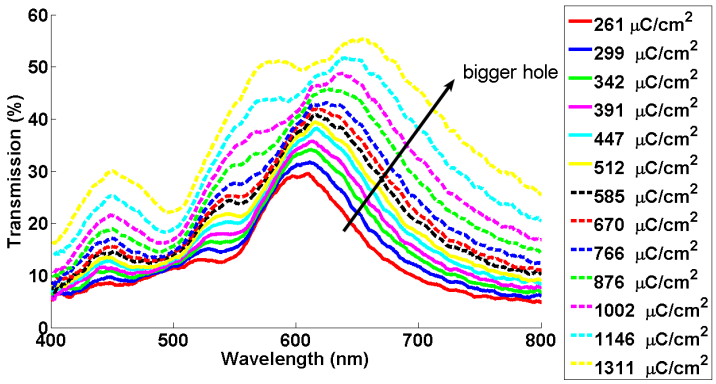
Figure 4.12: Plot of the simulated transmission of the three filters with parameters matching those in Table 4.2 (solid lines) compared to the transmission of the same filters with the radius of 50% of the holes is reduced by 10 nm (dashed lines).

mask hole size but varying the EB dose during lithography. This resulted in the higher EB dosage providing larger diameter holes. Notably in the red filter (Figure 4.13(b)) the secondary peak (λ= 440 nm) becomes more visible as the transmission increases.

Other parameters can also contribute to the variation in the transmission spectra; variations in the top SiO_2 layer have been shown to alter performance [108]. The transmission properties of the filters also depend on the material re-



(a)



(b)

Figure 4.13: The transmission of the (a) blue and (b) red filter for various electron beam doses (larger dose = bigger hole).

fractive index and absorption. The properties utilised in the simulation analysis are carefully chosen from literature [110]. Whilst these properties are determined for materials deposited in a similar manner, there nevertheless may be a small discrepancy due to unavoidable differences in equipment and process details. This could result in variation from the experimental material properties and could account for some of the discrepancy between the experimental and simulation results. These factors could be readily accounted for in the well-controlled conditions of a production foundry utilising dedicated equipment for each process step. Similarly further optimisation of the fabrication process would serve to reduce such variations. Further to this, if desired, such fabrication tolerances could be accounted for in the design process in an attempt to minimise the variation

in the structure performance, although such rigid fabrication tolerances are not necessary here, where the aim is to prove the concept of the device.

The simulations were conducted at normal incidence and no beam divergence. This zero divergence beam cannot be matched with the current measurement set up. The halogen lamp source provides incident light over a range of incident angles. Similarly the large numerical aperture objective (NA=0.5) results in an acceptance angle of 30° [32]. Due to the dependence of the magnitude of the wave vector on the incident angle, and therefore SPP resonant wavelength dependence on the incident angle (Equation 4.12), the SPP modes would be excited over a broader wavelength range. This would alter the transmission properties of the filters such that their peaks appear broader than expected. The plasmonic filters demonstrated an improved match to the CMF than that of current dye based filters (an example spectra of such a filter set is shown in Figure 2.3), and therefore can be implemented in trichromatic imaging systems to reduce the error in colour reproduction. Additionally, being metal based filters, they can be incorporated into the lower metal interconnect layers of a standard CMOS imager, or alternatively on top of the silicon of a BSI imager as discussed in Section 2.6.1.1 (lithography permitting). This would reduce the optical crosstalk between pixels. The ability to fabricate three filters in a single lithography process reduces the number of process cycles currently required in the fabrication of filters. An additional advantage of the plasmonic filters is their negligible transmission in the infra-red spectrum. This would reduce future system cost and complexity by removing the need for an additional IR blocking filter. These are currently required in digital imaging systems where the silicon based detectors are sensitive to transmission in this spectral range.

4.7 Chapter Summary

The colour sensitivity of current imaging systems is difficult to tune because it is limited by the material properties of the dyes used in the filters. For this reason it is difficult to improve colour reproduction of such systems. The greater flexibility in tuning the spectral response of plasmonic filters, by variation of the geometrical parameters via lithographic tuning, provides an alternative filtering technique

that can be applied to improve the colour reproduction of imaging systems. An automated design procedure using GA-based optimisation was implemented to take advantage of this and specifically optimise such structures, with multiple design parameters, to match a desired criteria. This was applied successfully to design filters matching the CMF. The simulated filters had FWHM of 81, 75 and 61 nm, for the red, green and blue filters respectively. This is a difference of only $3 - 25$ nm from that of the corresponding CMF. The resulting filter designs have been demonstrated experimentally. The fabricated filters had FWHM of 117, 150 and 100 nm and transmission of $25 - 35\%$. The materials used are fully compatible with the CMOS process therefore they could be implemented without complication. Such filters cannot be achieved using standard dye based filters currently applied in CMOS imaging. The plasmonic CMF can be applied in standard trichromatic imaging systems to reduce the colour reproduction error.

Chapter 5

Fabry-Pérot Filters

5.1 Introduction

For most consumer applications of *rgb* digital cameras, such as family photos, trichromatic imaging is sufficient to meet the colour reproduction needs of the user. The small variations in colour reproduction are visually acceptable to an untrained observer. For other applications, higher accuracy imaging and colour reproduction is required; for example, in vegetation assessment, knowledge of the reflection spectra, or the exact colour of the crop, is required. This provides information such as the nitrogen content or identifies when the crop is ready for harvest [11, 63, 64]. Another example is the accurate reproduction of artwork; artwork often is damaged or fades overtime, high accuracy digital imaging is used for recording, preservation and restoration throughout the world of art [6, 7, 8]. For applications such as these, multispectral imaging is required.

A detector containing more than three filters, with a narrower wavelength range pass band, can provide a better resolution representation of the spectrum of the incident image than that which can be achieved with *rgb* imaging. The common absorption filters currently implemented in imagers, and also the plasmonic filters presented in Chapter 4, have not been successful in demonstrating narrowband (FWHM$<$ 50 nm) filtering. Alternative techniques have been presented in literature to achieve this (Section 2.6.2) however these are not easily integrated within CMOS imagers. In this Chapter an alternative solution to pro-

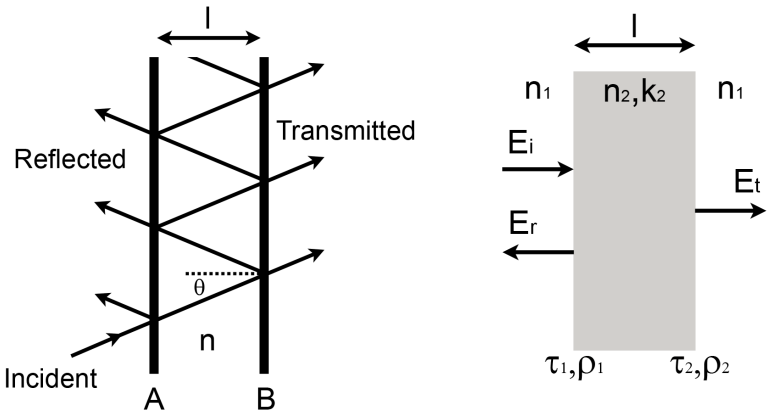
vide narrowband filters for application in CMOS imagers, using a metal-dielectric Fabry-Pérot (FP) resonator, is presented.

5.2 The Fabry-Pérot Resonator

The FP resonator has multiple applications, most notably in laser technology, to measure or control specific wavelengths of the incident light. It consists of two partially transmissive mirrors (this formation is also referred to as an etalon) or a dielectric with two reflective surfaces. Usually, but not always, the mirrors are plane-parallel [111]. The incident light is split into multiple components by reflections at the interface, and these reflections interfere with each other. Consequently transmission only occurs for select passbands of the incident light. The passband can be predicted by a straight-forward analysis; an overview is given here.

Light incident on the FP resonator is partially transmitted through the first mirror. As shown in Figure 5.1(a), multiple reflections occur between the two surfaces, here labelled A and B, for monochromatic incident light of wavelength λ. The light incident on the second surface (B) will also be partially transmitted at each reflection. Constructive interference will occur if the difference in the optical path length, at successive reflections from mirror B, is such that the rays are in phase. This will result in a transmission maximum. Destructive interference occurs for rays where this condition is not met, resulting in reduced transmission. The result is resonant filtering of the incident light.

Here, for ease, the transmission and reflection from the mirror-cavity interfaces alone shall be considered; ignoring the absorption present due to the transmission through the metallic mirrors. As previously detailed in Section 3.1.2 the transmission coefficients of multiple interfaces, with multiple reflections between the interfaces, can be calculated using the T-matrix method. The incident wave arrives from the left medium and there is no wave incident from the right. For simplicity the assumption is made that the medium to the left and right of the structure are the same, with refractive index n_1, and the dielectric cavity has refractive index n_2 (see Figure 5.1(b)). k_2 is the propagation wave number and l is the cavity length. The resulting electric fields from a dielectric slab are shown



(a) (b)

Figure 5.1: Diagram of the FP cavity; (a) the multiple reflections between the two surfaces A and B for a monochromatic light and (b) the electric field components of the cavity.

in Figure 5.1(b). Detailing the incident and reflected electric field (E_i and E_r respectively) in terms of the transmitted electric field (E_t), the following expression is derived:

$$\begin{pmatrix} E_i \\ E_r \end{pmatrix} = \frac{1}{\tau_1} \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} e^{ik_2 l} & 0 \\ 0 & e^{-ik_2 l} \end{pmatrix} \frac{1}{\tau_2} \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix} \begin{pmatrix} E_t \\ 0 \end{pmatrix} \tag{5.1}$$

Equation 5.2 is formed by multiplying out the matrix factors, which can then be solved to obtain the transmission response in Equation 5.3.

$$E_i = \frac{e^{ik_2 l}}{\tau_1 \tau_2} \left(1 + \rho_1 \rho_2 e^{-2ik_2 l}\right) E_t \tag{5.2}$$

$$\mathcal{T} = \frac{E_t}{E_i} = \frac{\tau_1 \tau_2 e^{-ik_2 l}}{1 + \rho_1 \rho_2 e^{-2ik_2 l}} \tag{5.3}$$

Here the T-matrix components are defined in terms of the reflection and transmission coefficients (τ and ρ respectively) to provide Equation 5.3 in the format commonly cited in literature [112].

$$T = \frac{n_1}{n_1} |\mathcal{T}|^2 = \frac{\tau_1^2 \tau_2^2}{1 + \rho_1^2 \rho_2^2 + 2\rho_1 \rho_2 \cos 2k_2 l}. \tag{5.4}$$

Where

$$\rho_1 = \frac{n_2 - n_1}{n_2 + n_1}, \quad \rho_2 = \frac{n_1 - n_2}{n_1 + n_2}, \quad \tau_1 = 1 + \rho_1, \quad \tau_2 = 1 + \rho_2. \tag{5.5}$$

This can be further simplified by using Equation 3.12 and the observation that, for this case where the medium to the left and right are identical, $\rho_1 = -\rho_2$.

$$T = \frac{\tau_1^2 \tau_2^2}{1 + R^2 - 2R\cos\delta} = \frac{\tau_1^2 \tau_2^2}{(1-R)^2 + 4R\sin^2\frac{\delta}{2}}. \tag{5.6}$$

Here δ, the phase delay between each of the successive transmitted rays[1] , is equal to

$$\delta = \frac{4\pi n_2 l \cos\theta}{\lambda}. \tag{5.7}$$

From Equation 5.6 it can be identified that maximum transmission occurs when $\delta = m2\pi$, where the order of interference m is an integer. From this the wavelength of the transmission peak can be defined by

$$m\lambda = 2n_2 l\cos\theta. \tag{5.8}$$

In other words, transmission peaks occur when the difference in the optical path length ($nl\cos\theta$) between each transmitted ray is an integer multiple of the wavelength. This relation demonstrates that the wavelength at which the transmission maximum occurs is dependent on l and n of the cavity, and the angle of incidence at the interface (θ). Traditionally, when applied as a filter, the FP cavity transmission peak is varied by adjusting l.

The spacing between the transmission maxima ($\Delta\lambda$) is referred to as the free spectral range (FSR). This can be calculated by

$$\Delta\lambda = \frac{\lambda^2}{2nl\cos\theta + \lambda}. \tag{5.9}$$

This is also a function of l and n. This is an important measure for this application as it is undesirable to have more than one transmission peak per filter within the visible band. A figure of merit for a FP filter is given by the ratio of the FSR to

[1] not to be confused with δ_d and δ_m as previously used in Figure 4.1(b) to represent the penetration depth

the FWHM ($\delta\lambda$)[1]. This fraction is known as the finesse ($\mathcal{F}$) and is a measure of ability of the FP resonator to resolve closely spaced spectral lines.

$$\mathcal{F} = \frac{\Delta\lambda}{\delta\lambda} \approx \frac{\pi\sqrt{R}}{1-R}. \tag{5.10}$$

As R increases[2], $\mathcal{F}$ increases, the transmission maxima become narrower and the FP resonator has a finer wavelength resolution.

As previously described, the dielectric cavity of the FP filter is enclosed by mirrors of a semi-transparent metal layer on either side. This increases the reflectivity of the interfaces, but also increase ohmic loss. Equation 3.11 is modified to account for the loss becoming:

$$R + T + A = 1, \tag{5.11}$$

where A is the fraction of the incident light absorbed in the metal. The inclusion of the mirrors therefore results in a reduction in the intensity of the transmission peaks relative to an unmirrored cavity of the same material. Additionally the phase change on reflection is no longer zero. This can be accounted for by including the phase change on internal reflection in the δ term, which will result in an increase in the optical thickness of the cavity. The additional phase change is typically sufficiently small in relative terms that it can be neglected [32]. When the films on the two sides of the cavity are identical, R is interpreted as the reflectivity for internal reflection and therefore Equation 5.6 holds. The reflection from a metallic surface is further explored in literature [112]. Equation 5.6 and 5.7 provide an estimate of the wavelength of the transmission spectrum maxima and are regularly applied for such FP filter structures. For a more thorough theoretical analysis, the T-matrix method can be applied. If the complex refractive index of the metallic mirrors are included, this can derive the transmission spectrum including absorption. It should be noted that the fabricated device mirrors will not be ideal; they will not be completely planar and will have surface roughness. This will result in increased FWHM and reduced transmission of the resonance

[1] not to be confused with δ in Equation 5.7

[2] R cannot exceed unity

peaks from that predicted analytically [113]. Such non-ideal mirrors however still provide suitable performance to enable practical application for filtering, and the performance of the fabricated structures can be accurately predicted theoretically using an ideal mirror model.

As discussed previously the position of the transmission peak of such filter devices is traditionally tuned by variation of l. If this approach was implemented in an array, within a CMOS process, it would require multiple lithography cycles to achieve. An alternative is to vary the dielectric cavity n. This approach has been explored previously in air filled cavities by variation of the pressure. Exploiting the fact that that air pressure proportional to $n - 1$. In this work effective medium theory (EMT) shall instead be applied to achieve a variation of n. By introducing artificial structuring into the cavity, at dimensions smaller than the optical wavelength, the macroscopic cavity properties can be altered. Such a nanostructured cavity can be treated as a homogeneous material with an effective refractive index.

5.3 Effective Medium Theory

EMT is applied to derive an analytical model of the macroscopic properties of a medium from its components. Here it is applied to quantify the effective refractive index n_{eff} of a nanostructured cavity, from the knowledge of the materials and dimensions. As light propagates through the cavity containing subwavelength structures it behaves the same as if the cavity was a homogeneous medium. Therefore with the application of EMT in the analysis of such a cavity, it is possible to represent the cavity as a homogeneous layer with equivalent properties. The structure therefore can be modelled as an equivalent layer within the T-matrix method, or the analysis procedure outlined in Section 5.2.

Both Born *et al* [114] and Rytov [115] present a derivation of the effective properties of media stratified in one dimension. The structure explored consisted of alternate homogeneous layers of two materials, with relative permittivity of ε_1 and ε_2, and a period Λ as shown in Figure 5.2. Rytov conducted an analysis of the field average over Λ. From the solutions of the averaged electromagnetic field the zeroth order effective permittivities were derived. For the TE and TM electric

field arrangement (propagating in the direction parallel to the layers), these are

$$\varepsilon_{TE}^{(0)} = \varepsilon_1 f + \varepsilon_2 (1 - f), \tag{5.12}$$

$$\varepsilon_{TM}^{(0)} = \left[\frac{f}{\varepsilon_1} + \frac{1 - f}{\varepsilon_2} \right]^{-1}, \tag{5.13}$$

where f is the fill factor of each material. For all mediums explored in this work it can be assumed that the permeability is that of free space ($\mu_r = 1$). For this reason this property, as used in the equations of [115], is not included.

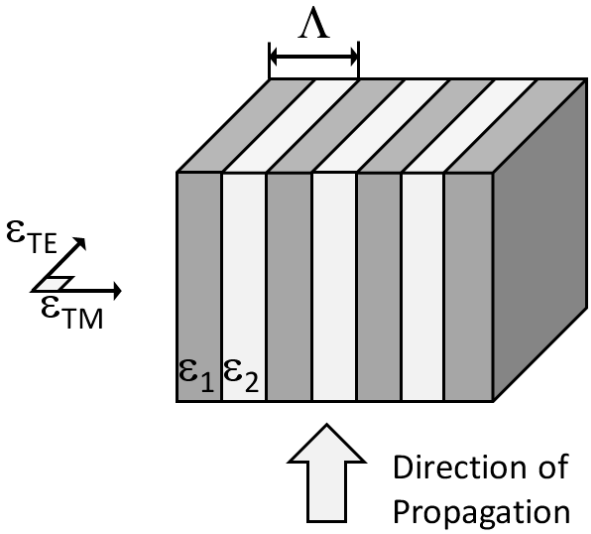


Figure 5.2: A schematic of the stratified medium as analysed by Rytov [115].

For greater accuracy in modelling larger period subwavelength structures, a correction factor can be included within the zeroth order EMT [115], yielding the second order EMT;

$$\varepsilon_{TE}^{(2)} = \varepsilon_{TE}^{(0)} \left[1 + \frac{1}{3\varepsilon_{TE}^{(0)}} \left(\pi(\varepsilon_1 - \varepsilon_2) \frac{\Lambda}{\lambda_0} f(1 - f) \right)^2 \right], \tag{5.14}$$

$$\varepsilon_{TM}^{(2)} = \varepsilon_{TM}^{(0)} \left[1 + \frac{\varepsilon_{TE}^{(0)}}{3} \left(\pi \frac{\varepsilon_{TM}^{(0)} (\varepsilon_1 - \varepsilon_2)}{\varepsilon_2 \varepsilon_1} \frac{\Lambda}{\lambda_0} f(1 - f) \right)^2 \right]. \tag{5.15}$$

Rytov identifies the limit on his EMT approach as $k\Lambda|n| \ll 1$. This requirement on the periodicity arises from the assumption that the field must be slowly

varying if a homogeneous material to give a suitable model of the structure. Others state that a limit of Λ less than $\frac{\lambda_0}{4}$ would be sufficient for the structure to be treated as a continuous film [116]. It shall be demonstrated that for the structures presented in this work, where Rytovs limit is exceeded, that a second order EMT evaluation of the structure is valid.

From Equations 5.14 and 5.15, it can be seen that the effective permittivity, and therefore effective refractive index (n_{eff}), of the structure is dependent on the polarisation state of the incident illumination relative to the surface periodicity. For the structures used in this work, which are periodic in two dimensions and homogeneous in the direction of propagation, an average of the two polarisation contributions is applied (Equation 5.16):

$$n_{\text{eff}} = \sqrt{\frac{\varepsilon_{TE}^{(2)} + \varepsilon_{TM}^{(2)}}{2}}. \tag{5.16}$$

5.4 Fabry-Pérot Filter with Nanostructures

In this work an alternative approach to the design and fabrication of FP filters is taken, whereby subwavelength structures are incorporated into the cavity of an FP filter for the first time. By altering the dimensions of the cavity structures, the effective refractive index of the cavity changes and the transmission peak of the filter shifts. Two variations are explored. First an FP cavity, with cylinders etched through the length of the stack, stopping at the lower mirror (note, the second design is explored in Section 5.5). The cavity n_{eff} is varied by adjusting the diameter of the cylinders (d) and therefore the air fill factor (f). A schematic of the structure is shown in Figure 5.3. The filter had upper and lower reflecting layers of a thin film of aluminium, and a nonstoichiometric silicon nitride (Si_xN_y) cavity. The higher index of the Si_xN_y, in contrast to the unity index of air holes, provided a suitable index ratio. These materials provided a suitable resonance tuning range in addition to the required optical transparency. The approach can be used with other geometries, e.g. pillars, to similar effect. A top layer of SiO_2 was also introduced into the design to act as an ARL. This design enabled the production of multiple filters in a single lithography cycle, reducing the fabrication

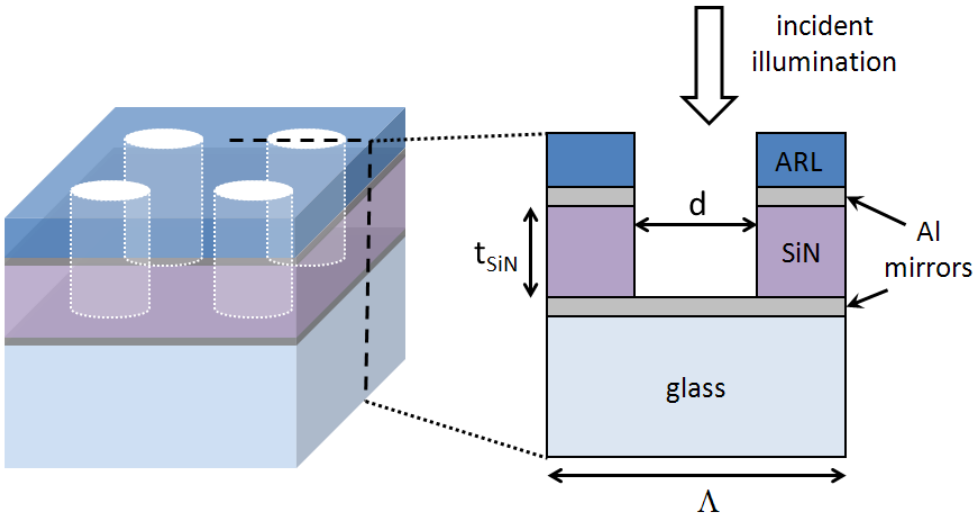


Figure 5.3: A schematic and cross-section of the Fabry-Pérot filter with etched air holes.

burden from that by variation of the cavity thickness alone.

The period of the etched holes (Λ) was selected as 200 nm because this met the condition of a subwavelength structure for the wavelength range of interest. The diameter of the etched holes was changed to provide different values of f. Here f is the fill factor of the etched air hole and is calculated by

$$f = \frac{\pi(d/2)^2}{\Lambda^2}. \tag{5.17}$$

5.4.1 Simulation

EMT theory, as presented in Section 5.3, can be applied to derive n_{eff} for the nanostructured cavity. The validity of this approach is confirmed by an analysis comparing the EMT-derived index to that of the cavity n_{eff} calculated from a FDTD simulation (TEMPEST). This was achieved by conducting a FDTD simulation of the cavity, as shown in Figure 5.4, and extracting the resulting transmission output. This was compared to the transmission spectrum of a T-matrix analysis of a single homogeneous layer, of the same t_{SiN}, for a series of n_{eff} values. The transmission responses of the two approaches are shown Figure 5.4(a). The refractive index values used were 2.05 and 1, for Si_xN_y and air respectively. At

a range of points, in the wavelength range of interest, the T-matrix transmission plot that most closely matched the FDTD simulation was identified. From this the n_{eff}, as used in the T-matrix analysis, was known. This value was representative of the refractive index of the FDTD simulated nanostructured cavity. The resulting dispersive n_{eff} deduced for this specific cavity ($t_{\text{SiN}} = 150\,\text{nm}, d = 160\,\text{nm}$) is shown in Figure 5.4(b). As can be seen this matched the n_{eff} value as calculated from EMT (Equations 5.13-5.17). At points of the transmission plot where the two lines cross-over, erroneous n_{eff} values were selected. At these sections it is the point of crossover which is selected as the nearest point, as opposed to the desired nearest point. These values therefore were removed during evaluation and the data averaged over 15 data samples to reduce the impact of this on the final calculation. Figure 5.5 plots the results of the analysis for a range of values of t_{SiN} and d (and therefore f). This demonstrates that second order EMT can be applied to deduce the n_{eff} of a nanostructured cavity for the range of cavity and hole dimensions used in this work. The transmission spectrum of the complete structure was predicted using the TEMPEST FDTD software. A cross-section

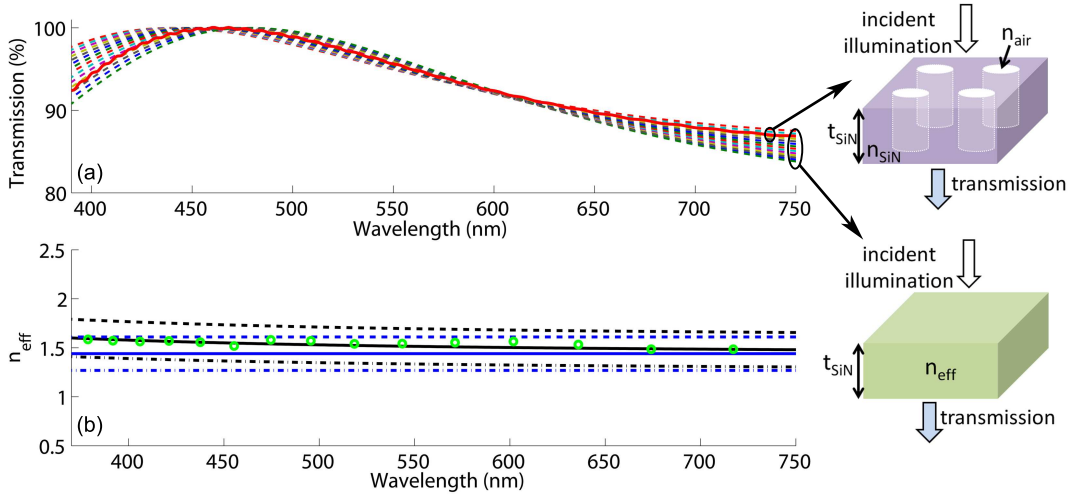


Figure 5.4: Analysis of the n_{eff} of the cavity; (a) a FDTD simulation of the cavity with nanostructures (solid) and T-matrix simulation of a solid cavity of various n_{eff} values ($--$). (b) The n_{eff} of the cavity calculated from first (blue) and second (black) order EMT theory. The TE ($--$), TM ($-\cdot-$) and averaged (solid) components are plotted. This is compared to the n_{eff} value ($\circ$) calculated from plot (a).

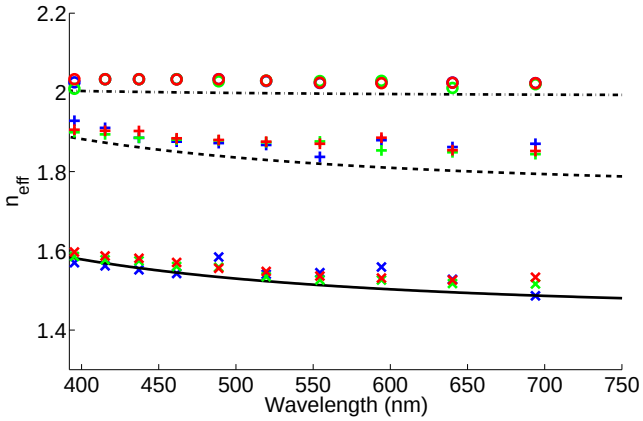


Figure 5.5: A plot of the n_{eff} of the nanostructured cavity as evaluated from Equations 5.13-5.17 for $d = 40$ ($-\cdot-$), 100 ($--$) and 160 nm (solid) for a structure of $\Lambda = 200$ nm. Additionally the n_{eff} predicted by FDTD and T-matrix analysis is plotted for $t_{\text{SiN}} = 150$ (blue), 250 (green) and 400 nm (red), and $d = 40$ ($\circ$), 100 ($+$) and 160 nm ($\times$).

of the 3D simulation domain is shown in Figure 5.6. The simulation domain was bounded in the z-dimension by PML and in the x and y-dimensions by periodic boundaries, so as to replicate an infinite square array of cylinders. The cell size was $\Delta x, y, z = 10$ nm for the complete simulation domain. The Al mirrors' dispersive behaviour was derived from the Drude + two critical points (D2CP) model (Appendix A).

An exploration of the structure dimensions was conducted to identify a set of narrowband filters covering the visible spectrum. The mirror and ARL thickness selected were 20 nm and 100 nm respectively. This provided suitable transmission and out-of-band rejection. The t_{SiN} value was selected to provide suitable FSR; providing a single resonance within a bandgap covering the visible spectrum. A set of 16 narrowband filters was created using five different t_{SiN} values. This was achieved by using the first and second resonances of different cavity thicknesses. The filters with t_{SiN} of 220 and 290 nm utilised the second resonance of the FP structure, while the filters with t_{SiN} of 130, 140 and 150 nm utilise the first resonance of the structure. The resulting set of filters is shown in Figure 5.7. The FWHM of the filters varies from $19 - 77$ nm. These simulation results show this

design can produce a set of narrowband filters that spans the visible wavelength range.

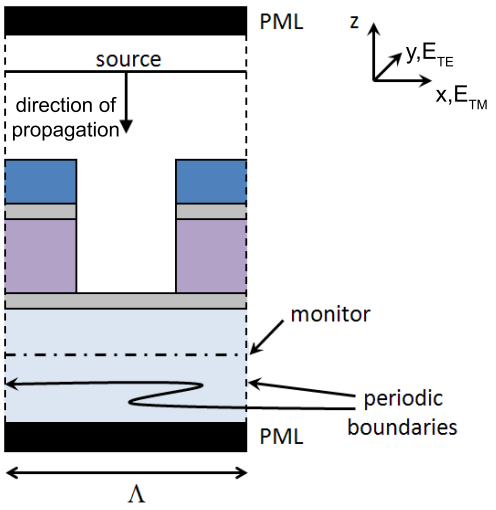


Figure 5.6: A schematic of the cross-section of the simulation domain of the FP filter with air holes.

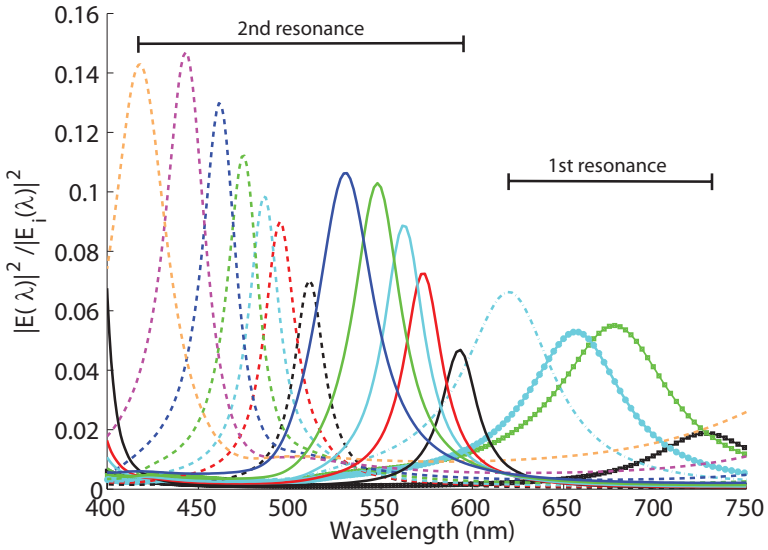


Figure 5.7: Simulated spectral response of a filter set with tuning achieved by variation of d. Filter $t_{\mathrm{SiN}} = 220$ ($--$), 260 (solid), 130 ($- \cdot -$), 140 ($- \circ -$) and 150 nm ($-\square-$). Filter d values are no hole (black), 60 (red), 80 (cyan), 100 (green), 120 (blue), 140 (magenta) and 160 nm (orange).

5.4.2 Fabrication

A nanostructured FP design was fabricated on a glass microscope slide. First the stack of materials was deposited. From the bottom to the top, the structure consisted of an Al mirror (20 nm), Si_xN_y cavity (200 nm), Al mirror (20 nm) and SiO_2 ARL (200 nm) as shown in the first stage of the fabrication diagram in Figure 5.8. The metal layers were deposited using a Plassys MEB 400S EB evaporator, and the dielectric layers by plasma-enhanced chemical vapour deposition. A layer of poly(methyl methacrylate) (PMMA) electron beam resist was then spin-coated on to the sample and exposed using the EB lithography tool. After development a layer of NiCr was deposited and a lift-off process performed to form a metallic etch mask. Finally, the sample was etched in CHF_3/Ar, $SiCl_4$ and CHF_3/O_2 in sequence for patterning the SiO_2, top Al and the Si_xN_y layer respectively. An interferometer was used to monitor the etch depth. The holes were etched through the structure, stopping at the bottom Al mirror as shown in Figure 5.8. For this structure, a square lattice of holes, with period $\Lambda = 200\,\text{nm}$, and a mask hole diameter of 100 nm was used. So as to to form multiple filters, several $50 \times 50\,\mu\text{m}$ arrays of holes were created. The EB dose was varied between the arrays to provide a variation in d, and therefore f.

5.4.3 Measurement and Analysis

The transmission spectra of the fabricated filters, was measured using a TFProbe MSP300 spectrometer. The results are shown in Figure 5.9(a). Here the filters are identified by the EB dosage labelled on the sample. The experimental structure demonstrated a single resonance and a FSR covering the visible spectrum. The fill factor of the holes is unknown, as the etch profile within the hole cannot be seen. A shift in the resonance to smaller wavelengths however is observed ($501 \rightarrow 479\,\text{nm}$) for filters with decreasing labelled EB dosage. This shift in the resonance towards shorter wavelengths results from the increase in fill factor as defined by the hole diameter during lithography. The FWHM of the filters shown are $60 - 64\,\text{nm}$. The resulting hole profiles varied across the array due to the non-optimised etch procedure, as shown in Figure 5.9(b). Better hole definition could be achieved with further development of the etch procedure, but

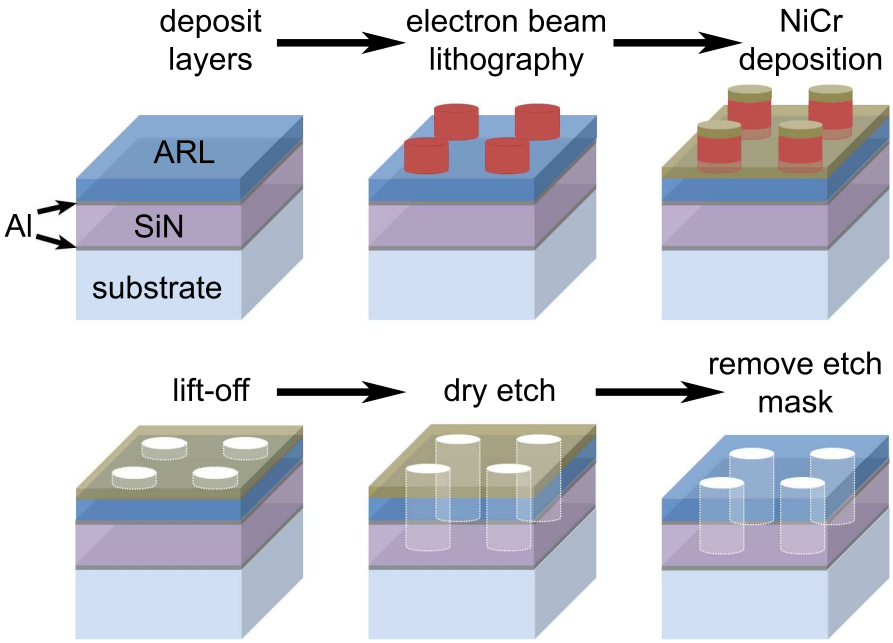
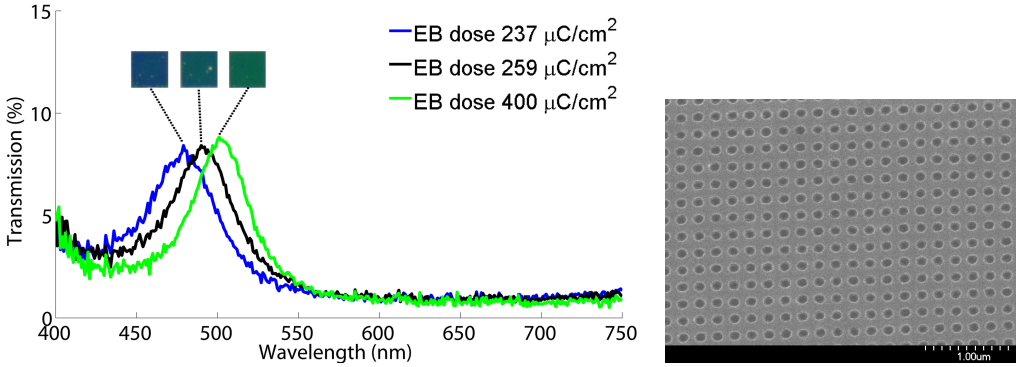


Figure 5.8: Diagram of the process flow for the fabrication of the filters with air holes etched into the structure.

this would only be necessary if it was desired to routinely manufacture filters in a commercial environment; the developed technique suffices for a proof-of-principle device. These experimental measurements demonstrate the validity of the presented design, and also the EMT based approach, for the production of multiple narrowband filters within a full visible bandgap, in a single lithography cycle.

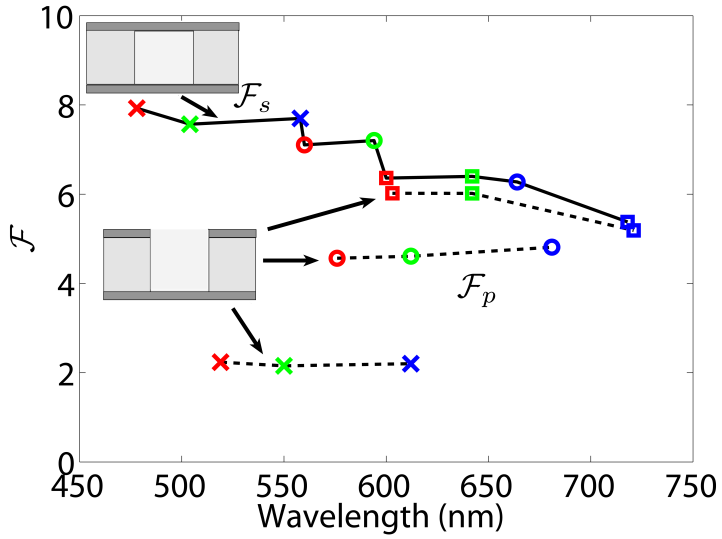
The non-monotonic shift in the wavelength of the resonance, seen in the filters spanning the entire visible range in Figure 5.7, is a consequence of the variation in reflectivity at the cavity-mirror interface. The change in refractive index of the Al across the spectral range, in addition to the reduction in n_{eff} with increasing hole diameter, varied the reflectivity and therefore the FWHM. Similarly the dispersive metal absorption resulted in variation of the filter maximum transmission. In addition to this reflectivity is also reduced due to the reduced area of the upper mirror with increasing size of the hole etched through the structure. This is further detailed in Figure 5.10(a), where the finesse, a measure of the FP cavity performance, as previously discussed in Section 5.2, for a range of different FP



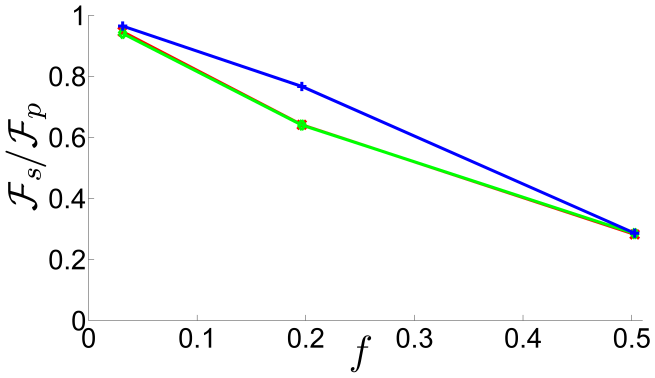
(a) (b)

Figure 5.9: The fabricated FP filters with etched air holes; (a) the transmission spectra of the filters, the insert shows a white light microscope image of the visible transmission of the filters, and (b) a scanning electron micrograph image of the top surface of a filter with $\Lambda = 200\,\text{nm}$.

structures is plotted. The solid line in the plot depicts the finesse achieved for a cavity with a completely intact, or solid, upper mirror ($\mathcal{F}_s$). As can be seen $\mathcal{F}$ is dispersive, reducing at longer wavelengths, due to the dispersion of the aluminium and cavity refractive index. For a structure with solid mirror mirrors this dispersive behaviour is consistent independent of f and t_{SiN}. The dashed lines in the graph plots the finesse calculated for a cavity where the upper mirror is perforated ($\mathcal{F}_p$), for three values of f. This shows a reduction from that achieved with a solid mirror FP structure. The dispersive behaviour is generally consistent between the two different designs, for various t_{SiN} cavities; for example comparing the square markers (indicating $f = 0.0314$), for different colours (representing the cavities of different t_{SiN}), between the $\mathcal{F}_p$ and $\mathcal{F}_p$ plots shows both structures demonstrate the same general trend due to the dispersion. The variation between the finesse for the two structures is, for cavities with the same f and t_{SiN}, due to the reduced reflectivity of the upper mirror with increasing f. From 5.10(b) it can be seen that ratio of the finesse ($\mathcal{F}_s/\mathcal{F}_p$) has a decreasing linear trend with increasing f. For $f = 0.5027$, or $d = 160\,\text{nm}$, the finesse is reduced to 2 or 72% of that achieved FP with a solid mirror. However for $f = 0.0314$, or $d = 40\,\text{nm}$, the reduction is only $\sim 5\%$. Since $\mathcal{F}_p$ and $\mathcal{F}_s$ are dispersive (while the ratio is not) the minimum $\mathcal{F}$ will be found at longer wavelengths.



(a)



(b)

Figure 5.10: A comparison of (a) the finesse of a Si_xN_y-air FP structure for both a complete and perforated mirror. For $f = 0.0314$ (□), 0.1963 (○), 0.5027 (×) and $t_{SiN} = 150$ (blue), 130 (green) and 120 nm (red). The finesse ratio, of the perforated and solid mirror structures, variation with f is also shown (b).

The reflectivity of the perforated mirror (R_p) can be derived if Equation 3.12 is further defined as $R_p = r_s r_p$. Where r_s and r_p are the reflection coefficients of the solid and perforated mirror respectively. A useful quantity is the coefficient of the finesse (F) which is defined as

$$F = \frac{4R}{(1-R)^2} = \left(\frac{1}{\sin(\frac{\pi}{2\mathcal{F}})}\right)^2. \tag{5.18}$$

If R_p is substituted into the Equation 5.18 to provide the finesse coefficient for a perforated mirror structure (F_p), this can be solved for r_p:

$$r_p = \frac{F_p + 2 - \sqrt{F_p + 1}}{F_p r_s}, \tag{5.19}$$

and similarly substituting in R_s and solving for r_s

$$r_s = \frac{F_s + 2 - \sqrt{F_s + 1}}{F_p r_s}. \tag{5.20}$$

F_s and F_p were extracted from the transmission output of a set of FDTD simulations of example FP structures, previously detailed in Figure 5.10. Using Equation 5.18 and 5.20, r_s was calculated from the solid-mirror filter simulations. An approximate linear fit was calculated, as plotted in Figure 5.11(a), for convenience using the 'fit' to a linear polynomial curve function in MATLAB. With knowledge of r_s, in addition to data extracted from the perforated mirror simulation analysis, r_p can be derived from Equation 5.19. Plotted in Figure 5.11(a) are the values of r_s, at the transmission peaks, for various cavity structures (the cavity dimensions were previously detailed in Figure 5.10). In Figure 5.11(b) the r_p value derived from the perforated cavities is plotted as a function of the hole diameter. From this it can be seen that the variation in r_p can be modelled by degrading the reflectivity of the solid mirror by the ratio of the metal to total area

$$r_p(\lambda) = r_s(\lambda)(1 - f). \tag{5.21}$$

This model is compared to data extracted from the simulations in Figure 5.11(b). The vertical spread of the data is due to the dispersion (as seen in Figure 5.11(a)), therefore the model is plotted for the maximum and minimum values of r_s. As can be seen there is good agreement between the model and data.

The model identified in Equation 5.21, combined with the theory presented in Section 5.2, can be applied to aid in the design of a nanostructured FP cavity filter set. For a single cavity the range over which filter resonance can be tuned is limited by the ratio of the refractive index.

$$\lambda_{\max} : \lambda_{\min} \frac{n_{\max}}{n_{\min}} \tag{5.22}$$

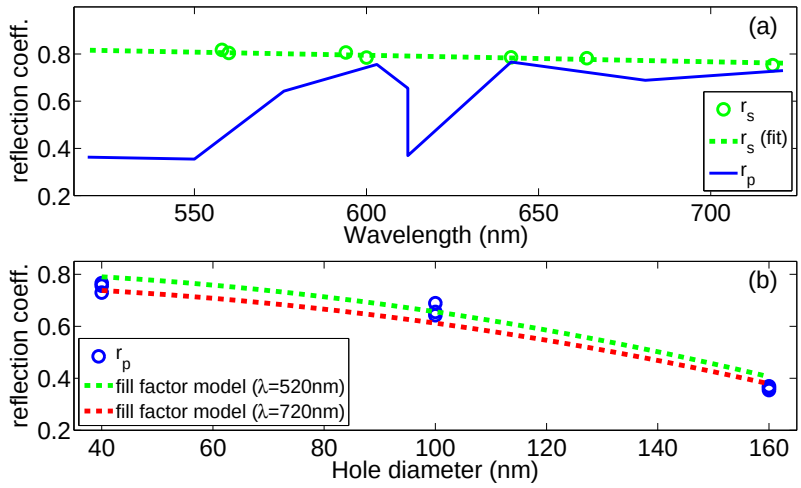


Figure 5.11: Calculation of the reflection coefficient r_p of a perforated mirror (a) and comparison to its prediction by degrading the reflection coefficient r_s of the solid mirror in accordance with the fill factor (b).

where for the structure used in this work;

$$n_{\text{ratio}} = \frac{n_{\text{max}}}{n_{\text{min}}} = \frac{n_{\text{SiN}}}{n_{\text{SiN}}(1 - f_{\text{max}}) + n_{\text{air}} f_{\text{max}}}. \tag{5.23}$$

The maximum desirable fill factor (f_{max}) can be estimated from the minimum acceptable finesse ratio (see Figure 5.10), as larger holes increase the tuning range achieved with a single cavity length but reduce (F). From this the number of cavity lengths, otherwise referred to as the number of filter sets (N_{sets}), required to cover a desired range (the range defined λ_{min} to λ_{max}) can be derived from

$$N_{\text{sets}} = \frac{\log(\lambda_{\text{max}}/\lambda_{\text{min}})}{\log(n_{\text{ratio}})} \tag{5.24}$$

A design guide for the structure is shown in Figure 5.12. This can be used to calculate the N_{sets} required for a given finesse ratio constraint, or vice versa. Equation 5.24 is plotted alongside the finesse ratio extracted from Figure 5.10 (inset). The dotted line indicates how to extract the required information from the graph; for example $N_{\text{sets}} = 5$ gives a finesse ratio of 0.66.

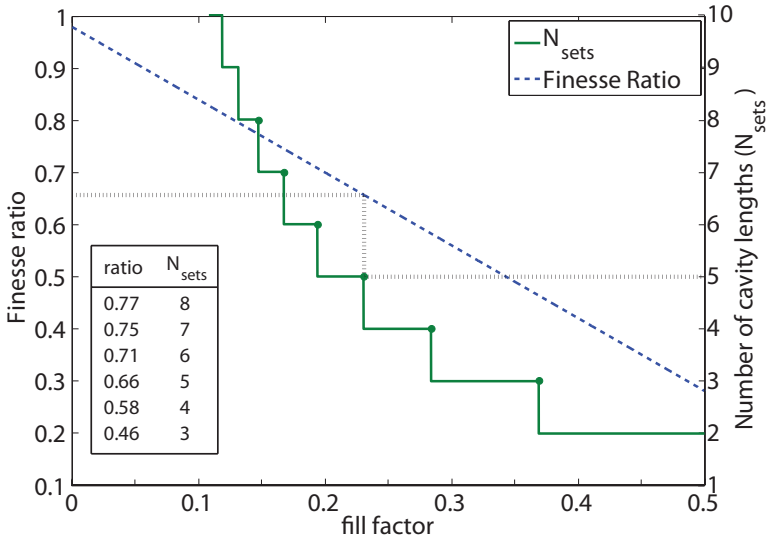


Figure 5.12: Design guide to calculate the number of filter sets for a minimum finesse ratio. The dotted line provides an example of its use.

5.5 Backfilled Fabry-Pérot Filter Design

As discussed for the nanostrucutred design presented in Section 5.4, the upper mirror area is reduced with increasing hole size. As a consequence R is reduced, and therefore the FWHM increased, with increasing f. One way to overcome this issue is to backfill the hole that is etched into the cavity before deposition of the upper mirror. This enables a fully intact mirror to be deposited and the full tuning range to be exploited without reduction in the filtering performance.

5.5.1 Structure

This design is similar to the structure presented in Section 5.4, here however the holes are backfilled and the cavity planarised before deposition of the upper mirror and ARL. This is shown in Figure 5.13. One filling material suitable for implementation within this design is poly(methyl methacrylate) (PMMA). This material is optically transparent and can be spin-cast onto the structure to fill the holes. PMMA has a refractive index of 1.49. Although larger than that of air, this still provides a suitable index contrast to the Si_xN_y cavity to allow tuning of the filter resonant peak.

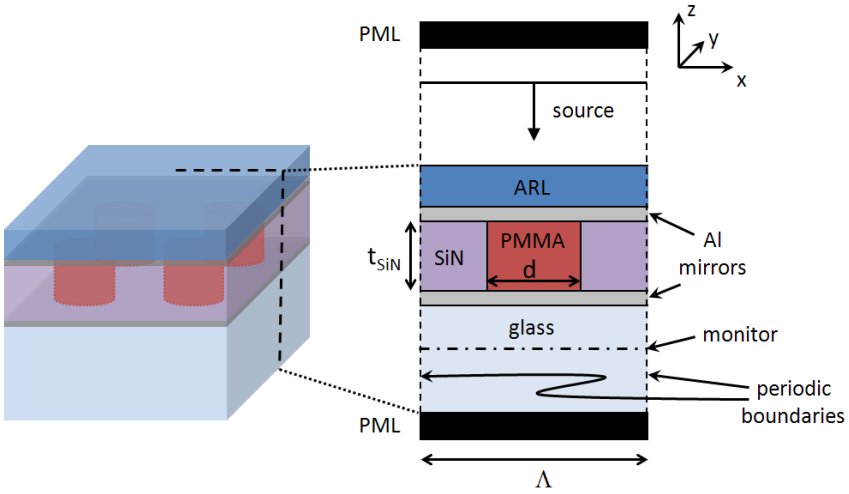


Figure 5.13: A schematic the Fabry-Pérot filter with PMMA filled holes and a cross-section of the simulation domain.

5.5.2 Simulation

Using FDTD simulations of a PMMA-Si_xN_y nanostructured cavity and the analysis process as detailed in Section 5.4.1 (previously applied to the air filled cavity), the n_{eff} of the cavity can be derived. The n_{eff} calculated, for a range of cavities with various t_{SiN} and f, is shown in Figure 5.14 (markers). The n_{eff} calculated from second order EMT (Equations 5.13-5.17) is plotted as lines, for various f. The insert of Figure 5.14 is a diagram of the top view of a period of the cavity; this demonstrates the nanostructuring in relation to each of the cavity f values plotted. Comparing the n_{eff} derived from the two different methods demonstrates that the n_{eff} of a PMMA-Si_xN_y nanostructured cavity can be modelled by second order EMT theory.

A study of the range of dimensions of the complete structure was conducted using TEMPEST FDTD simulation software. A cross-section of the simulation domain is shown in Figure 5.13. The Al mirrors are modelled by the Drude + 2 critical points model. The cell size was $\Delta x, y, z = 10\,\text{nm}$ across the whole simulation domain. The boundaries in the x and y dimensions were periodic so as to simulate an infinite square array of holes. In this analysis, the inverse structure were also explored: Si_xN_y pillars backfilled with PMMA. These could

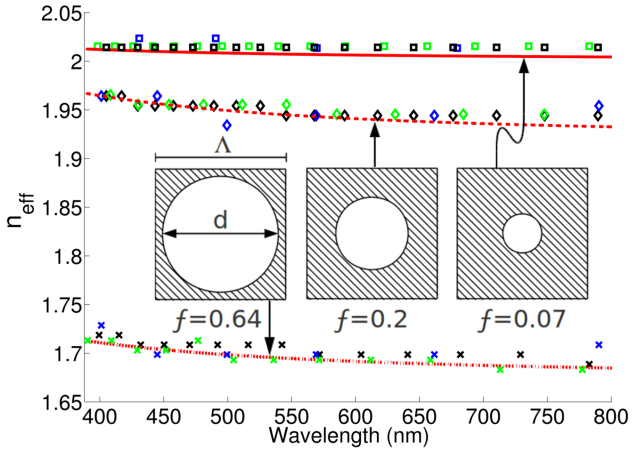


Figure 5.14: A plot of the cavity n_{eff} of the filled cavity, with $d = 60$ $(-)$, 100 $(--)$ and 180 nm $(\cdots)$ for $\Lambda = 200$ nm. This is compared to the n_{eff} calculated from FDTD simulation for $t_{\text{SiN}} = 800$ (black), 500 (green), 200 nm (blue). With $d = 60$ $(\square)$, 100 $(\Diamond)$ and 180 nm$(\times)$.)

be used in place of the larger-diameter holes where otherwise maintaining the etch profile may be difficult. In this simulation study, it was identified that a sufficient FSR, could be achieved using only two different cavity thicknesses. A sufficient FSR is one where each filter has a single transmission peak in the visible range. The resulting filter set is demonstrated in Figure 5.15. The filters use the first resonance of the design with $t_{\text{SiN}} = 150$ nm and the second resonance of the design with $t_{\text{SiN}} = 240$ nm, for the long and short wavelength halves of the band respectively. The filters shown use hole or pillar structures with $d = 60$ to 180 nm, $\Lambda = 200$ nm, and ARL and mirror thicknesses of 100 nm and 20 nm respectively. The FWHM of the filter set shown varied from $13 - 36$ nm, which corresponded to a finesse $\mathcal{F}$ of $10 - 30$.

The N_{sets} required to cover the visible range ($\lambda_{\text{min}} = 400$ nm, $\lambda_{\text{max}} = 750$ nm) can be calculated using Equation 5.24. For this new design, using $n_{\text{PMMA}} = 1.49$ and $n_{\text{SiN}} = 2.05$, it is calculated that the N_{sets} required to cover this range is 2. Similarly, the cross-over point of the two sets predicted from Equation 5.22 is approximately 545 nm. The simulation analysis therefore has good agreement with the theoretical prediction of the tuning range from the design equations.

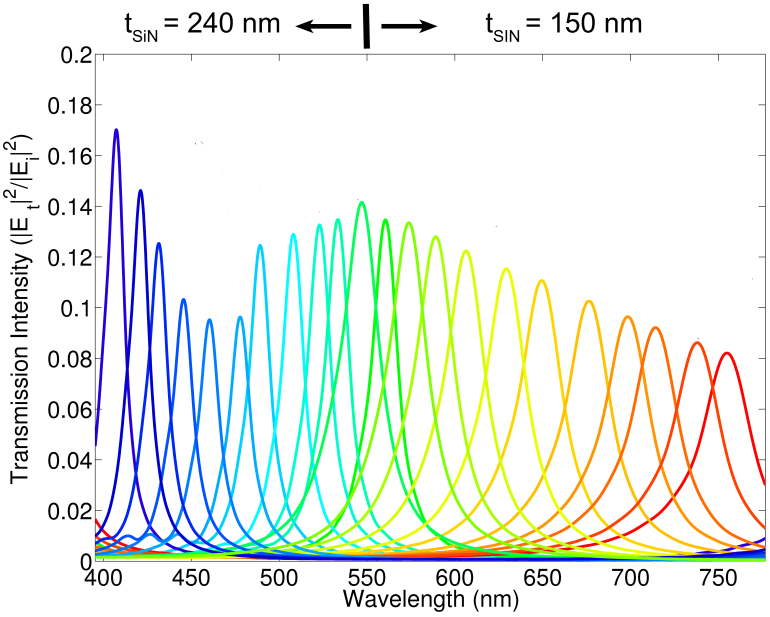


Figure 5.15: The FDTD simulated spectral response of a FP filter set using a Si_xN_y-PMMA nanostructured cavity using two different cavity thicknesses (t_{SiN}).

5.5.3 Fabrication

Six subsets of the filters were fabricated using a single cavity length $t_{SiN} = 200\,nm$. The fabrication process is outlined in Figure 5.16. The bottom mirror and the Si_xN_y cavity were deposited onto a microscope slide by a Plassys MEB 400S EB Evaporator and an Oxford Instruments PECVD tool respectively. A mask layer defined the multiple arrays of holes, each a different filter. The mask was spin-coated ZEP520A EB resist, pattered by a Vistec VB6 UHR EWF lithography tool, and developed in o-xylene. The holes were etched using CHF_3/O_2 in a Plasmalab 80+. After resist removal, a layer of approximately 100 nm thick PMMA was spin-coated onto the etched cavity to backfill the holes, and the structure was baked to harden the PMMA. In this proof of principle device the overfilled PMMA, as present after the spin on PMMA stage in Figure 5.16, was not polished away. The upper Al mirror was deposited and a further 100 nm layer of PMMA was spin-coated on top of the whole stack to form the ARL. The array of six filter subsets, each with a different mask hole size, was fabricated simultaneously by variation of the hole size defined in the lithography mask.

Different filter passbands were produced in each of the sets by further altering the EB dosage ($237 - 2000\,\mu\mathrm{C/cm^2}$), while maintaining the etch mask hole diameter, to achieve fine variation of the hole size.

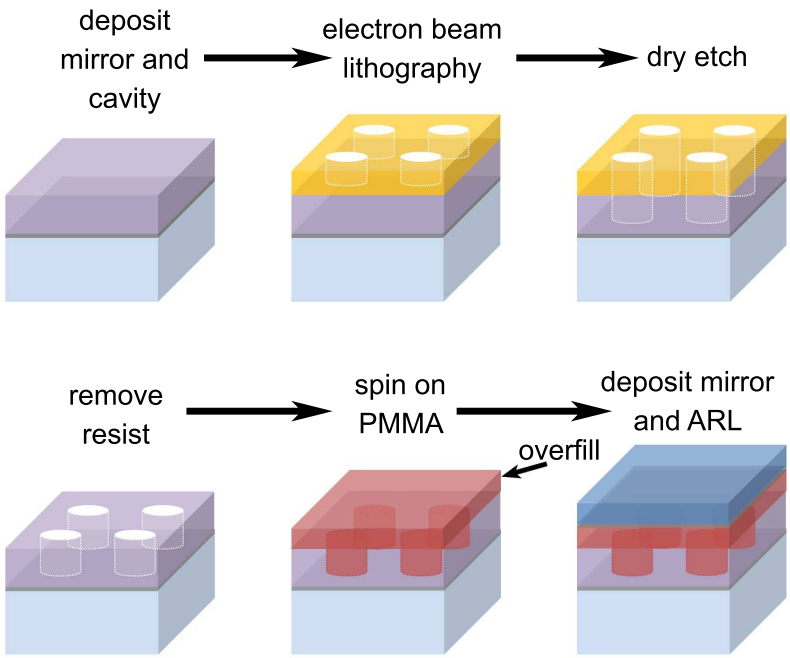
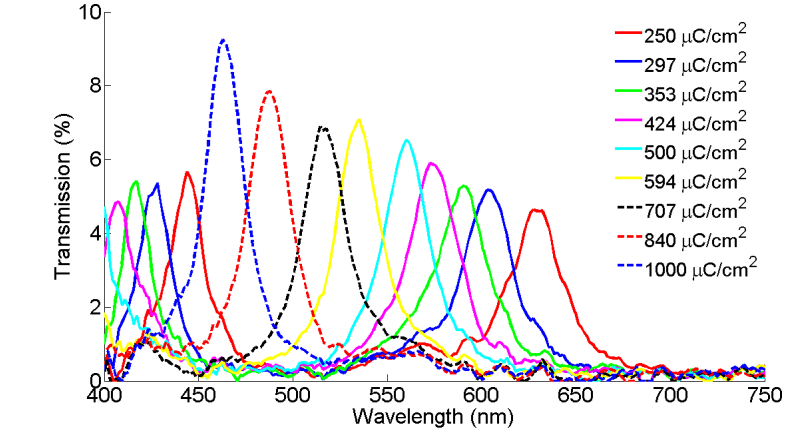


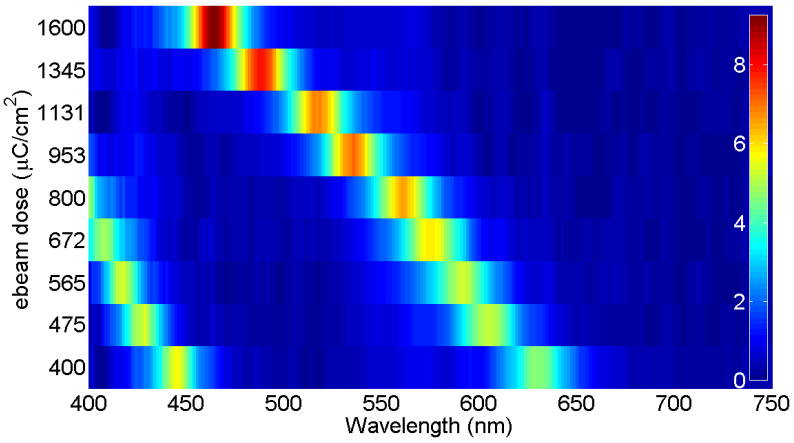
Figure 5.16: Fabrication procedure for the backfilled FP filters, where the PMMA overfill is not polished away.

5.5.4 Measurement and Analysis

The transmission of the six filters subset was measured using a TFProbe MSP300 spectrometer. A plot of the measured response, for an example set of filters with $d = 120\,\mathrm{nm}$ for various EB doses, is shown in Figure 5.17. It can be observed that the transmission peak shifted to shorter wavelengths for increasing EB dose, in this case representing an increase in f, as was expected. For some of the peaks there was an additional resonance within the visible band. This performance differs from that predicted in the simulation of the structure, even when accounting for the additional PMMA present on-top of the cavity after backfilling of the holes. The measured transmission performance demonstrated a larger effective cavity length (an increase in n_{eff} and/or t_{SiN}) from that of the simulation.



(a)



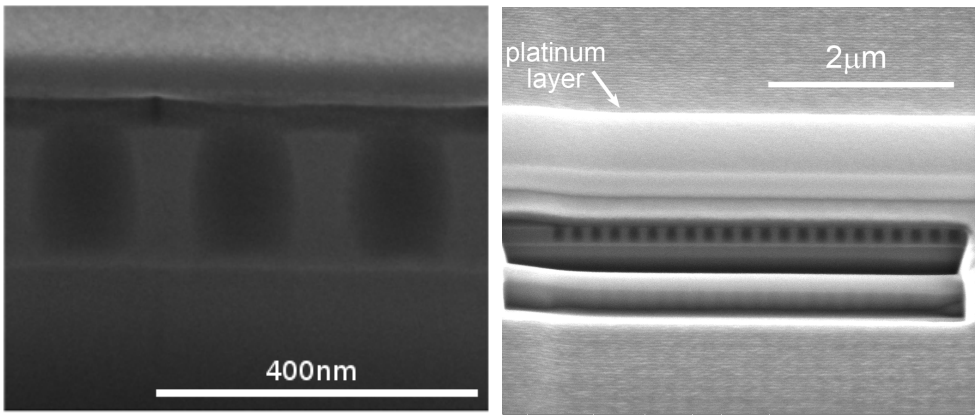
(b)

Figure 5.17: The transmission of a set of filters with $d = 120\,\text{nm}$ for various electron beam doses; (a) the measured spectral response of the filters where line colour indicate the EB dose and (b) a 2D plot of the same filter set where the colour indicates the transmission.

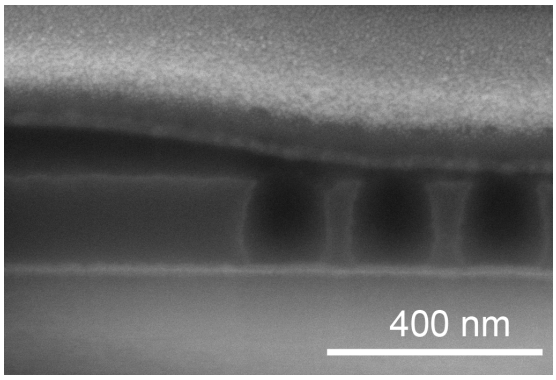
To understand this deviation in the transmission performance from that predicted in simulation, the fabricated structure was examined carefully. Using a focused-ion-beam (FIB), a hole was formed in one of the filters and a image of the cross-section was obtained using a scanning electron micrograph (SEM). The resulting images are shown in Figure 5.18. The additional layer present in 5.18(b) is a layer of platinum, deposited prior to the FIB procedure, to protect the structure from damage. From these images t_{SiN} was measured to be approximately

200 nm. An additional observation was that the thickness of the PMMA layer, left between the cavity and the upper mirror after backfilling, also varied. Filters with a larger hole size had a thinner layer of PMMA than those with a smaller hole due to the reduced amount of PMMA material left after infilling of the holes. This is identified most predominantly in Figure 5.18(c), an image taken at the edge of a filter where the structure transitions from a solid cavity to a nanostructured cavity. It can be seen that the thickness of the PMMA layer is reduced significantly in the nanostructured section due to the infilling of the holes. This PMMA layer and the structure dimensions extracted from the images however cannot alone account for the observed increase in the effective cavity length. This increase therefore must be the consequence of a higher refractive index material present in the cavity than previously assumed in the device simulations. The presence of a higher refractive index cavity material would indicate a silicon(Si)-rich Si_xN_y cavity, instead of a standard nonstoichiometric Si_xN_y cavity assumed in simulation. It has previously been demonstrated, with this specific deposition tool, that an increase in the Si content of the material can occur if the SiH_4 relative flow rate is increased during deposition [117]. Accounting for the Si-rich Si_xN_y cavity enabled the effective cavity length of the fabricated filters to be successfully modelled.

A T-matrix simulation was conducted of the complete structure pictured in Figure 5.18(a); including the glass substrate, ARL (85 nm), nanostructured cavity($t_{\mathrm{SiN}} = 200\,\mathrm{nm}$, $d = 140\,\mathrm{nm}$), Al mirrors (15 nm) and the 90 nm layer overfill of PMMA. Within the simulation the refractive index of the Si_xN_y was replaced by a refractive index derived from first order EMT of a Si and Si_xN_y mixture. It was identified that if the Si_xN_y material was replaced by a mixture of 60% Si- 40% Si_xN_y, the simulation performance had good agreement with the measured transmission. The result is plotted in Figure 5.19, also shown is the measured transmission from a T-matrix simulation of the filter with standard Si_xN_y ($n = 2.05$). This Si-rich Si_xN_y material present in the cavity resulted in a reduced FSR and transmission from that designed due to the increased refractive index and absorption of Si. This however also resulted in a greater n_{ratio} than anticipated, as a consequence of this the transmission peak of the cavity has a tuning range covering the full visible spectral wavelength range.



(a) (b)



(c)

Figure 5.18: Scanning electron micrograph images of the cross-section of a filter.

Using this cavity material, a full set of filters therefore can be fabricated in a single cavity length, reducing the fabrication demand from the design with a standard Si_xN_y cavity (where two cavity lengths are required to cover the visible range). This approach would require the incorporation of an addition highpass filter; to filter out the additional lower wavelength peak present in some of the filter transmission spectrum. A set of 23 narrowband filters covering the visible range were formed from the sets of fabricated filters, as shown in Figure 5.20(a).

The measurements were taken with a bandpass filter, with a cut-off wavelength of 525 nm, between the sample and the spectrometer objectives. The measured filters had a transmission of $4 - 13\%$ and a FWHM of $22 - 46$ nm across the range. This experimental result suggests non-standard materials could be usefully

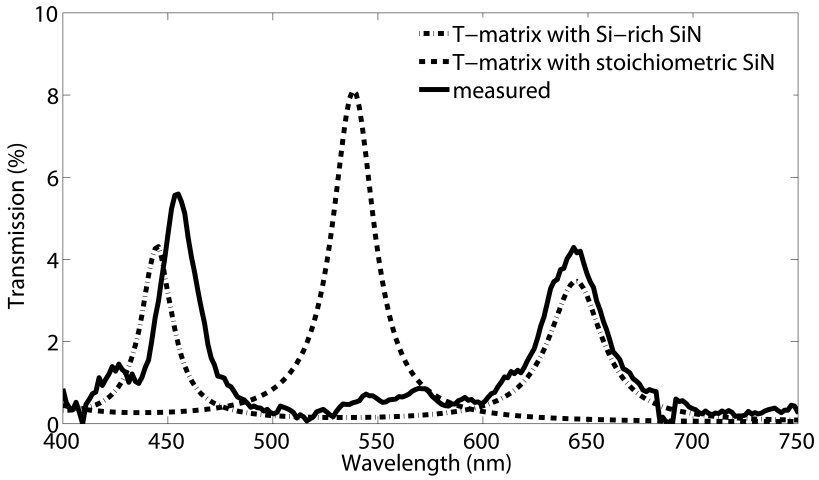


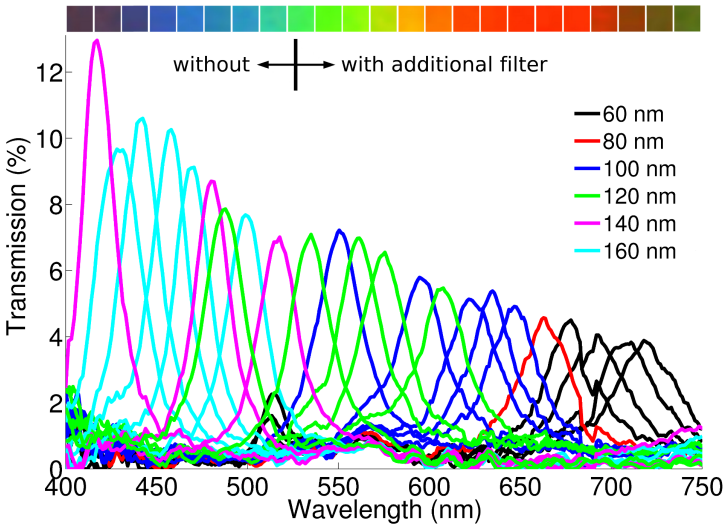
Figure 5.19: A comparison of the measured and simulated filter response for both PMMA-Si_xN_y and PMMA Si-rich(60%)Si_xN_y(40%)cavities.

exploited. A set of T-matrix simulations, using a Si-rich Si_xN_y cavity model, replicating the measured filters are also shown in Figure 5.20(b).

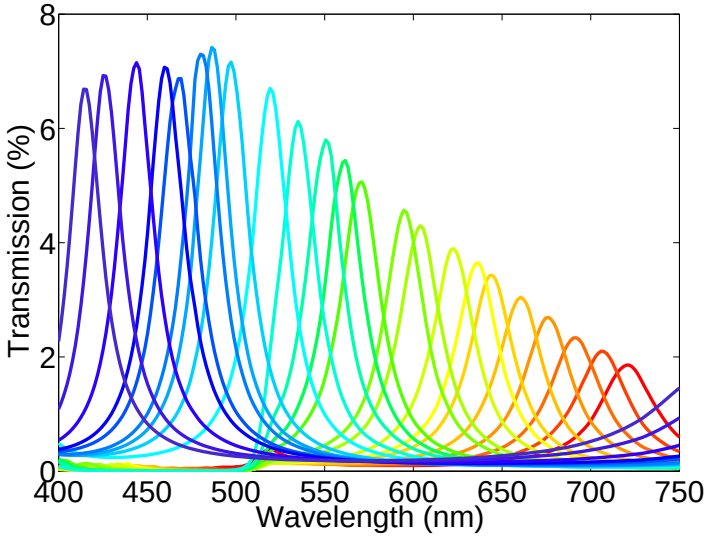
5.6 Chapter Summary

A set of narrowband filters covering the visible range has been designed, simulated, fabricated and characterised. The filter passband is tuned by altering the lateral dimensions of the nanostructures via the lithography mask. The effective refractive index of a FP filter is altered by the introduction of nanostructures into the cavity. This is easily achieved by CMOS processing techniques. This design enables the fabrication of multiple filters in a single cavity thickness, using a single lithography cycle. This reduces the fabrication demand from that of the traditional tuning method of varying the vertical cavity thickness. In addition the use of EMT for the evaluation of the effective refractive index of a nanostructured cavity is validated.

Two designs are presented; with and without backfilling of the nanostrucutres. Narrowband filters were simulated and demonstrated experimentally for both designs. For the structure without backfilling, a model quantifying the reduction in the reflectivity of the upper mirror with increasing hole size is derived. This can



(a)



(b)

Figure 5.20: A plot of (a) the spectral response of the fabricated set of filters and (b) the T-matrix simulation of the filters, with a Si-rich Si_xN_y cavity. In plot (a) the line colour indicate the nominal hole diameter in the lithography mask before EB dosage adjustment. A montage of the white-light microscope image of the transmission of the filter is shown (inset). The additional filter used in (a) is also applied in the simulation spectra in (b) for clarity.

be used in conjunction with EMT, and theory of FP cavity, to aid the design of filter sets using this structure dependent on the needs of the application. It was shown in simulation that a set of 16 narrowband filters covering the visible range can be achieved using only five cavity lengths. This is a significant reduction from the 16 required if variation of the cavity alone is applied; i.e. this significantly reduces the number of lithography steps required. Using backfilling of the nanostructures the upper mirror can be maintained intact, at the expense of additional fabrication process. Using this approach, a full set of filters covering the visible range can be achieved using two different Si_xN_y cavity lengths. Alternatively a single cavity length of a high-index material, and a bandpass filter, could be used. Using this approach a set of 23 narrowband filters were demonstrated; these were fabricated in a single lithography cycle.

Both these designs enable the fabrication of narrowband filter sets with a free spectral range covering the full visible range, at reduced fabrication cost and complexity than by variation of the physical cavity length alone. Such filters could be arrayed across an image sensor to enable high resolution simultaneous detection of the full visible spectrum in a compact system. Both approaches use only materials fully compatible with the CMOS process to illustrate that such filter sets are suitable for direct integration onto CMOS image sensors in industrial foundries.

Chapter 6

Application of Filters for Imaging

6.1 Introduction

This Chapter evaluates the performance of the filters from Chapters 4 and 5 in terms of image reproduction. Chapter 4 described plasmonic filters for red, green and blue imaging that better mimicked the human eye, while Chapter 5 described a narrowband multispectral filter set based on Fabry-Pérot cavities, offering 23 filters across the visible band. Since the filters have not been integrated within a camera, in order to see the image they would produce, the filters wavelength response must be applied in the virtual domain using test images for which the full spectral data is available. In order to enable the reproduction of a scene imaged through a set of filters, the device-dependent signals must be transformed into XYZ values (Section 2.2). First the conversion system must be trained; creating a transformation matrix (A) to enable mapping between the imager output (with scene spectral content $S_{\text{train}}(\lambda)$, illuminant $L(\lambda)$ and filter response $D(\lambda)$) and the desired CIEXYZ values for the imaged colour. Once the transformation matrix has been derived this can be applied for the translation of any scene, imaged through the same system, for reproduction for display or storage. Using the XYZ values derived from the imaging system and the CIEXYZ values expected for a colour, the reproduction error (ΔE_{2000}) for the system can be calculated. The training, reproduction and evaluation procedure, and the required input and output data, is outlined in the flow chart shown in Figure 6.1. This process

is applied to the filter sets presented in this work to demonstrate the colour reproduction and evaluate the reproduction error.

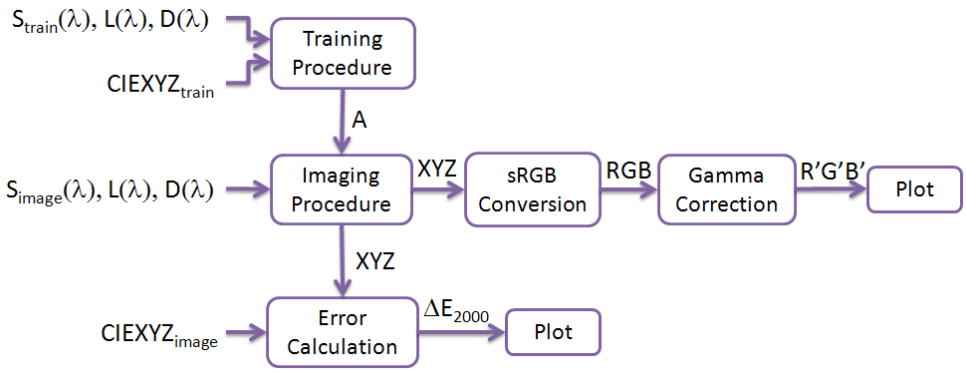


Figure 6.1: The image reproduction and evaluation process flow.

6.2 Colour Reproduction Procedure

This section describes the 'train-and-test' procedure that was implemented to provide a colour reproduction procedure, incorporating the proposed filter sets, for image reproduction and analysis. To begin, for each filter set, a training procedure was applied and from this the transformation matrix, A, was derived. This can be done computationally for this system as the imager response is represented by the filter spectra, and the spectra of the illuminant and colour patch are known. Manual imaging is not required. For an imaging system where the system response is unknown, A would be derived by a training procedure involving the manual imaging of each of the training patches. The imager response and the illuminant would be amalgamated with the colour spectrum within the imager channel response. For such systems additional processing or user input is required to extract the illuminant and imager response. The output of such a system is a digitised form of the rgb data as identified in Equation 2.3 or a 3 channel system. This data would be substituted into the same training procedure outlined in Figure 6.1 in place of the spectral data. For this analysis only the reproduction performance of the filters is to be examined therefore a representative

illuminant is selected and no imager response is included.

The procedure to translate the image system rgb information to device-independent XYZ values is outlined in Section 2.3.2.1. For this image reproduction procedure, the device-dependent signals were derived computationally using Equation 2.3. $L(\lambda)$ of standard illuminants for the visible range are freely available [118]. Here D65, the standard illuminant for northern sky daylight, was applied. The CCCR set of colour patches was used for training [24]. This provides a set of example colour patches ($S(\lambda)_{\text{train}}$) from across the colour range as shown in Figure 6.2. The filter spectra, measured using a spectrometer, were used as the imaging system response $D(\lambda)$. $D_{r/g/b}(\lambda)$ of the three channels designated r, g and b in Equation 2.3, were represented by the spectra of the red, green and blue simulation and measured spectra of the plasmonic filters shown in Figures 4.8 and 4.10 respectively. For the multispectral filters set there were 23 channels, with $D_{1-23}(\lambda)$ taken from the measured spectra of the FP filters as shown in Figure 5.20.

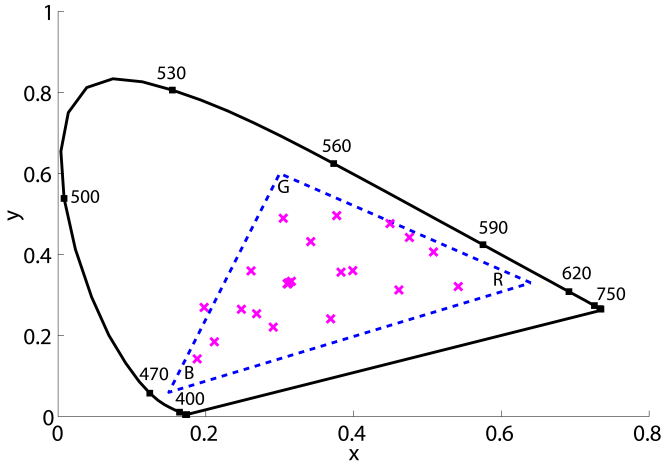


Figure 6.2: A plot of the CIE 1931 chromacity diagram, with markers at selected wavelengths (□), and the CCCR training patches in D65 illumination (×). Also plotted is the chromacity diagram of the sRGB colourspace (−−).

Using $S(\lambda)_{\text{train}}$, $L(\lambda)$ and the CMF, the CIEXYZ tristimulus values of each of the colour patches was derived from Equation 2.1. A least squares polynomial modelling process, using a second order polynomial as shown in Equation 2.5, was utilised to derive A. Once A has been derived for the filter set from the training

process, this transform can be applied to map any input spectrum ($S(\lambda)_{\text{image}}$) to a device independent XYZ value for storage or reproduction. As previously described this can again be replicated computationally using Equation 2.3 if the input spectral data ($S(\lambda)_{\text{image}}$, $D(\lambda)$ and $L(\lambda)$) are known. A is derived from the training procedure for the system and as before the system response $D(\lambda)$ is represented by the filter spectra.

The transformation of such recorded spectra from the device-dependent system to the device-independent XYZ representation can be evaluated. Using a set of test patches and A of the system, the XYZ value for each patch, as if imaged with the filters, is derived. The corresponding desired CIEXYZ value for each of the test patches is also calculated. Then using the colour error formulae, as previously outlined in Section 2.3.3, to compare the values, the magnitude of the error from the reproduction using the particular filter set can be evaluated. The CIEDE2000 formulae is computationally complex therefore the software used here to calculate the error values was initially verified using the example test CIELAB input values and corresponding ΔE_{00} values provided in [28].

6.3 Colour Reproduction Performance

The 'train-and-test' procedure was applied to evaluate the two plasmonic and FP filter sets for the reproduction of colour. The performance of these filter sets is compared to that achieved with an example three channel trichromatic dye and photonic crystal filter set. The example filter spectra used were extracted from graphs presented in [16, 58, 60]. An identical procedure was applied to all the filter sets; A for each system was identified and the error in the reproduction of the Munsell Book of Colour set of 1269 patches was calculated. All input spectra were in the range $400-750\,\text{nm}$, at $5\,\text{nm}$ wavelength intervals. The resulting error calculated is detailed in Table 6.1.

To visually demonstrate the reproduction error a random selection of Munsell patches and an example imaged scene, as reproduced using procedure outlined for the different filter sets, are shown in Figure 6.3. For each colour patch, the XYZ values (derived from the test procedure) were transformed to sRGB for display. This process is detailed in Section 2.3.2.3. In Figure 6.3, the patch reproduction

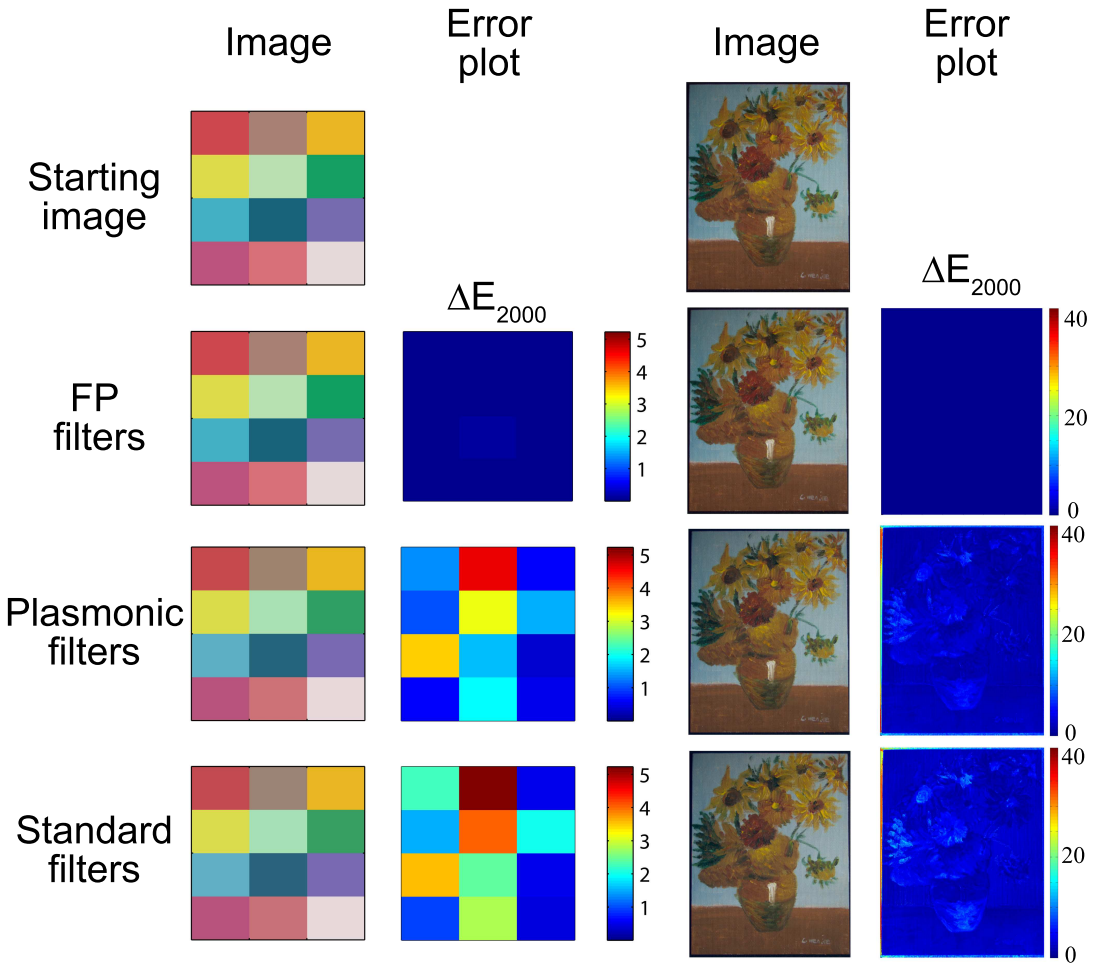


Figure 6.3: A plot of the resulting sRGB colour for a random selection of Munsell patches. Also plotted is the error measured for each patch compared to the CIEXYZ (top left). The blue error plots represents almost no error and hence the best colour reproduction. For the standard filter set the filter spectra from [16] was used.

Table 6.1: Colour reproduction error of the Munsell colour patch set for a selection of filters, where a lower error value represents a more accurate reproduction of the imaged colour.

Filter Set	$\mathbf{\Delta E_{ab}}$			$\mathbf{\Delta E_{00}}$		
	max	min	mean	max	min	mean
My 23 FP (measured)	0.348	0.003	0.063	0.343	0.002	0.047
My Plasmonic (simulation)	10.831	0.059	1.870	6.986	0.046	1.413
My Plasmonic (measured)	6.392	0.160	2.985	9.223	0.152	2.222
Typical imager RGB filters [16]	26.045	0.219	3.637	13.592	0.169	2.711
Multi-layer FP RGB filters [58]	8.694	0.102	1.993	16.206	0.202	2.714
E field tunable sensor [60]	27.693	0.472	5.747	12.850	0.237	4.027

using the filters presented in this thesis are compared to that of a RGB dye filter set from a CMOS imager, referred to as the standard filter set. The filter spectra used for this is extracted from [16] and are plotted in Figure 2.3. The top row of images is the sRGB plot of the desired image derived from the CIEXYZ values. The plots below this are the same image, reproduced from a system incorporating the different filter sets. It should be noted that all colour images presented in this work are in sRGB colour space therefore are best viewed using a medium with the ability to replicate this colour space.[1]

The ΔE_{00} value calculated from the XYZ values and the desired CIEXYZ for each of the filter sets is plotted in the second column. The two right-hand columns are the result of the same process applied to a complete image. The example image used was obtained from [119]; the scene was imaged with a multispectral camera and the spectral data for each pixel of the scene between $400 - 700\,\text{nm}$ in $10\,\text{nm}$ steps is provided. Using this spectral data the test procedure can be applied to each pixel and the image reproduced from the XYZ values derived. In comparing the ΔE_{00} plots for both the array of colour patches and the example scene, for the 3 channel filters systems, on average the error is lower for the plasmonic filters; i.e. the plasmonic filters give better colour reproduction than the standard RGB filters. This was the original intention behind designing the plasmonic filters to mimic the human eye response. This assessment, using the

[1] A computer monitor would provide a suitable display for viewing (sRGB is an approximation of the colour gamut of the most common computer monitors). Conversion to alternative colour spaces, for example for printing, may result in variation in the reproduced colour.

measured error and by visual comparing the reproduced colour, corroborates the analysis presented in Table 6.1. The average ΔE_{00} calculated for the image of the painting shown in Figure 6.3 was 2.5602 and 3.2818 for the plasmonic and standard filters respectively. This analysis of the colour error confirms that as the filter spectra of the plasmonic filters are a closer match to the CMF response than standard dye filters, the application of such filters in an imaging system would enable a reduction in error in the reproduction of a scene using this CIE system.

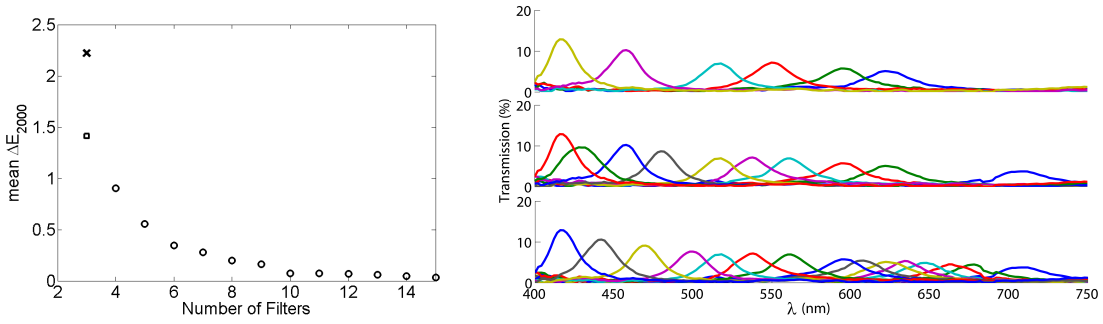
For the FP filters the average ΔE_{00} calculated for the image of the painting is 0.0695. The all-blue error plots in Figure 6.3 demonstrates that there is almost no error in the reproduction of the starting image. This is expected because its greater number of imaging channels would give better colour reproduction. This analysis confirms the significant reduction in colour error, compared to a trichromatic system, that can be achieved with the application of a multispectral imager. This demonstrates the advantage of applying such a filter set in applications where minimising colour reproduction error is important.

6.4 Analysis of Filter Set Size

For imaging applications where good reproduction of colour is the priority, as opposed to examination of a particular spectral band or reproduction of the spectra, acceptable performance can be obtained with less than 23 filters. A GA procedure was designed to identify a set of filters from the 23 filter fabricated, as shown in Figure 5.20, that would provide the minimum mean ΔE_{00}.

The number of filters allowed in the set is defined by the user at initiation, and the members of the set are selected from the 23 available FP filters. The GA used the mutation and crossover processes as described in Section 3.2.1. The settings of $p_c = 0.6$, $p_m = 0.05$ and 100 generations was used, the same parameters as identified in Section 3.2.2. The fitness of each trial set of filters in the population was defined by the mean ΔE_{00} error in the reproduction of the Munsell patches. The error was calculated using the same 'train-and-test' procedure as applied in the previous section. No individual filter could be selected twice within a filter set. The process was repeated to identify optimum filter combinations for a set

size of 6 to 15 filters. The minimum error, achieved by the optimum filter set, for each set size is plotted in Figure 6.4(a) alongside that achieved with the simulated and fabricated plasmonic filter sets. The filter spectra of the filters for the 6, 10 and 15 filter sets, as identified by the GA procedure, are plotted in Figure 6.4(b). The training and reproduction procedure was applied to reproduce two images using the filter sets presented in Figure 6.4(b). The resulting colour image and plot of the ΔE_{00} value is shown in Figure 6.5. As can be seen by these images the filter set using only 6 filters has the largest error, with a maximum ΔE_{00}= 9. The mean ΔE_{00} calculated were 1.0411, 0.2368 and 0.1087 for the image of the beads on the left, and 1.3579, 0.2719 and 0.2298 for the image containing the toys on the right. This was for the 6, 10 and 15 filter sets respectively. The mean and maximum error is reduced when increasing the numbers of filters in the set. This analysis using a real scene corroborates that observed in the plot of the mean error from the reproduction of the Munsell patches, as shown in Figure 6.4(a). In that analysis it can be seen that for the optimum sets identified by the GA, that the average error reduces with the increasing number of filters in a set.



(a) (b)

Figure 6.4: The results of the GA optimisation procedure (a) the minimum mean error of the GA (∘), fabricated plasmonic (×) and simulated plasmonic (□) filter sets, and (b) the filter spectra of the selected set of 6, 10 and 15 filters.

From this it can be identified that using a set of 6 FP filters significantly reduces the colour reproduction error from that achieved using trichromatic imaging. The error can be further reduced by increasing the number of filters, the number of filters therefore can be selected dependent on the colour reproduction requirements of the application. In the plot in Figure 6.4(a), the improvement

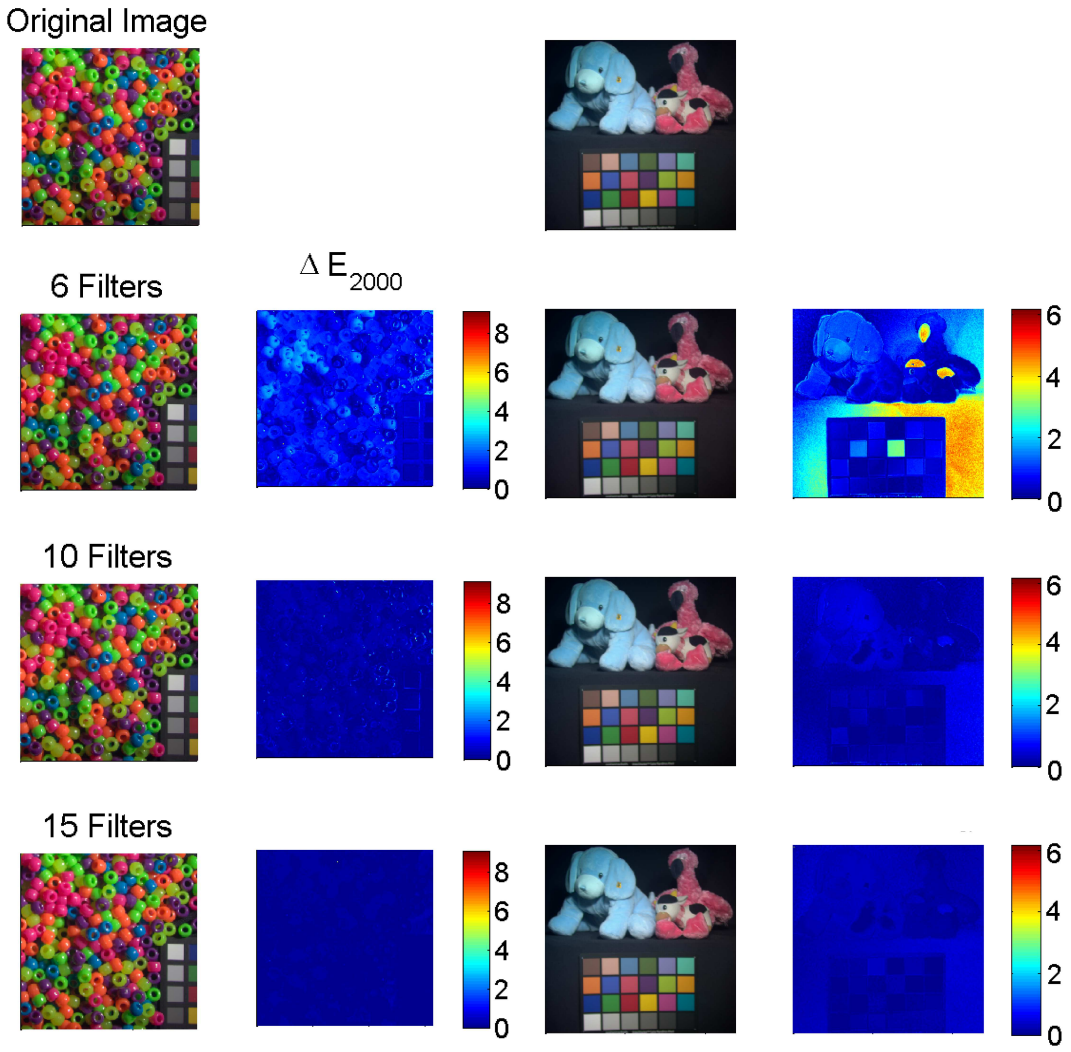


Figure 6.5: The reproduction of two different scenes if imaged using different multispectral filter sets, as identified from the GA search, using the training procedure outlined. Also plotted is the error measured for each pixel of the scene compared to the CIEXYZ (top image).

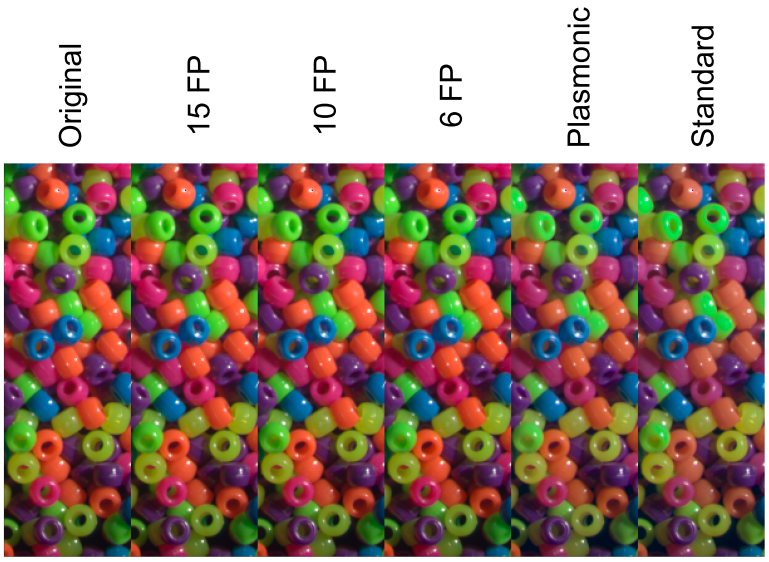


Figure 6.6: The reproduction of a section of a scene for different filter sets for easy comparison of the colour reproduction in reference to the original image.

in colour error becomes less significant after the set reaches 10 filters. For applications requiring the reproduction of the spectrum, or examination of specific wavelength bands of interest, larger filter sets may be required than that required for colour reproduction. The visible colour reproduction performance of each of the filter sets presented in this Chapter can be compared in Figure 6.6; here a section of an image has been reproduced for each of the filter sets using an identical training procedure. The resulting colour images have been plotted adjacent to each other, using this the reproduction can easily be compared to the original image, on the far left section of the plot, and the other filter sets. Now that the colour reproduction performance of the new filter sets have been compared, next a more detailed example of an application shall be considered, in which the reduced reproduction error capabilities of the presented narrowband multispectral filter sets can be exploited.

6.5 Example Application of a Multispectral Filter Set

Multispectral imaging is of most interest for applications where such colour reproduction error, or alternatively the exact reproduction of the colour, is significant. One example is distinguishing colours that are similar to each other, for example in military applications to distinguish a camouflaged vehicle from the surrounding scenery. Another application is in the food industry where multispectral imagers can be applied to detect the variations of colour in fruit such as bananas, mandarins and tomatoes [10, 11, 64]. The reflectance spectra of such fruit is determined by the absorbing compounds such as pigments, water and dry matter. The chemical changes that take place during ripening cause colour change. Traditionally, such judgements on fruit are determined by human observation, using standard colour charts. For a familiar example, the shift in the colour of a banana skin from green to yellow indicates ripening of the banana. Another example is the detection of the early stages of damage in fruit, this is difficult to identify casually by the human eye. A multispectral imaging system would enable the identification and quantification of minor colour variations and would increase the speed and accuracy of food processing. They could be applied as a non-destructive system to identify when the fruit should be picked, or for monitoring of the quality during ripening or processing.

An example of the application of the multispectral filter sets presented in this work is presented in Figure 6.7. For this example a general application is assumed and therefore the previously described training procedure, including the complete colour range provided by the CCCR chart, is applied. For more specific applications the training procedure could be modified and an optimum filter set identified for the application. For example, to detect the riping of bananas, training could be conducted using only the green and yellow training patches. In this example four patches are explored, they include two types of tree and two examples of green paint on different surfaces. The spectra are freely available [120] with accompanying descriptions. Patch 1 is described as ‘deciduous trees’, 2 as ‘conifer trees’, 3 as ‘a smooth, slightly translucent coat of alkyd gloss paint on construction lumber (pine wood)’ and 4 as ‘a single coat of olive green alkyd

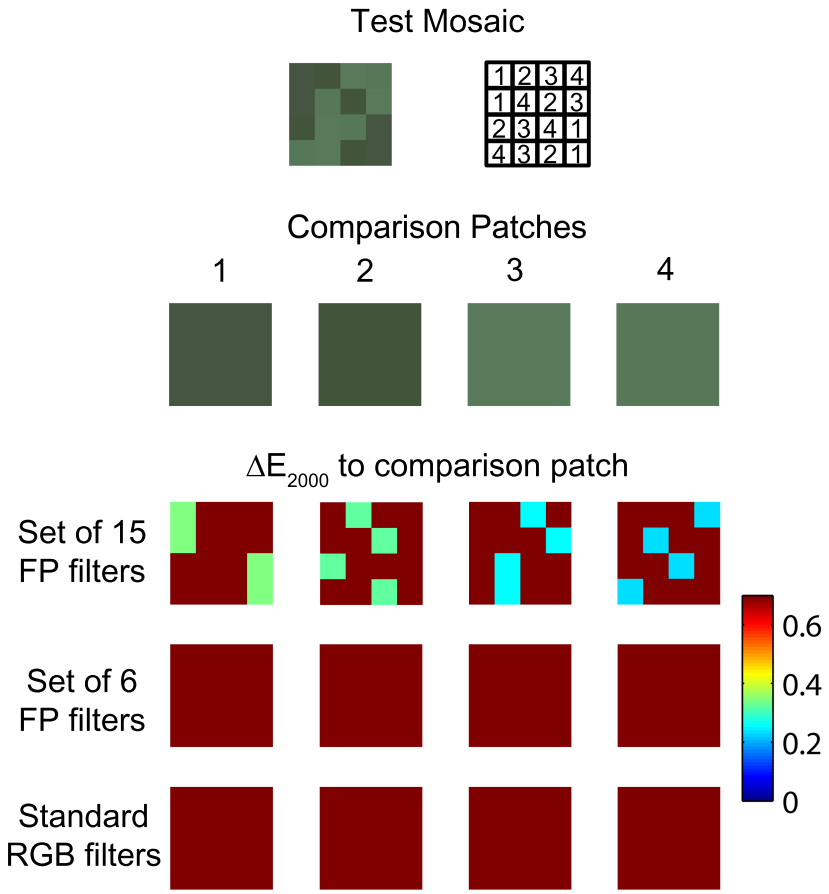


Figure 6.7: Example of the identification of visually similar colours using a multispectral imager. The error between each comparison patch and the array is plotted for several filter sets. The colour scale is limited to a maximum of 0.7, and all errors above this are plotted the same colour. This provides a colour space for identification of the similar patches in the array. Note the correct patches can only be identified using the 15 filter set.

gloss paint, on an aluminium surface'. The sRGB reproduction of the spectra, or the 'comparison patches' as otherwise referred, are shown in the second row of Figure 6.7. The ΔE_{00} difference, calculated from the CIEXYZ values, between patches 1 and 2 was 2.4324, and between 3 and 4 was 0.7656. The measured ΔE_{ab} was 3.9805 and 0.8869 respectively. This evaluation corresponds with the guideline provided in Section 2.3.3 as it can be seen that for a standard observer it is difficult to distinguish between patches 3 and 4. The colour difference between

patches 1 and 2 is discernible on close inspection of the patches next to each other. An array of the patches was created, as shown in the top row of the image, and this was compared to a single comparison patch. The ΔE_{00} value calculated for each patch in the array, when compared to the selected comparison patches, was plotted in the column below each patch, for each filter system reproduction. This is applied with the aim to identify each instance of the comparison colour patch in the array. This is achieved by plotting the ΔE_{00} value as shown in the bottom three rows of Figure 6.7. If the scale is reduced so any patch with an ΔE_{00} value over 0.7 is plotted as red, therefore a false colour map imposed, this can then be used as an image space to identify patches with ΔE_{00} less than this. In other words the squares of the array that match the comparison patch will be plotted as the representative colour from the colour map. All other patches, with an $\Delta E_{00} \geqslant 0.7$ will be plotted the same red colour, this therefore enables the easy identification of the matching patches within the array. As can be seen in Figure 6.7, using this process only the 15 channel FP filter set could identify the colour patches from the array. Even the patches that are difficult to distinguish in sRGB colour space are easily identified. Each imaged patch will have an error in reproduction due to the imaging and reproduction process, this error should be sufficiently small so that other patches of similar colour have a ΔE_{00} larger than that of the reproduction error. In this particular demonstration exploring the three filter sets this can only be achieved in a system with 15 channels.

This simple example demonstrates the application of multispectral imaging systems beyond standard image reproduction. With the filter design as presented in this work enabling the simplified fabrication of FP filters in an array, the improved colour reproduction afforded by multispectral imagers could become more accessible. Such systems could be applied where previously the application of a standard spectrometer was deemed costly or impractical.

Chapter 7

Conclusions and Future Work

In this Thesis, I designed, simulated, measured and evaluated two different visible range filtering techniques for use on digital imaging chips. These have been developed in order to improve the colour filtering performance of a CMOS imager, compared to that achieved with current BEOL filter technology. The main achievements of this study are:

- I developed a fully automated optimisation procedure for the design of plasmonic filters with arbitrary transmission spectra. This combined a genetic algorithm with a FDTD simulation tool to explore the large search space of design parameters to identify an optimum design.
- I applied the automated optimisation procedure to the design of a set of plasmonic filters with a transmission spectra matching the CIE colour matching functions (1931). These functions represent the colour response of the human eye and the application of such filters would enable improved colour reproduction. This was also the first demonstration of a single filter with a double peak as present in the $\bar{x}(\lambda)$ function.
- I was the first person to combine EMT and FP filters to create a narrowband filter set suitable for fabrication in an array on a digital imaging chip. The key benefit of my design is that it requires far fewer lithography steps than conventional filters.

- I demonstrated that even with a perforated upper mirror, a FP cavity could operate successfully as a narrowband filter. To evaluate the performance of such filters I presented a theoretical model quantifying the reduction in reflectivity with reduction in the mirror area. This can be applied to aid in the analytical design of such filters for a minimum acceptable reflectivity performance.

- I applied the plasmonic and backfilled FP filter sets within an image reproduction process to predict the performance of the filters in imaging applications. This process demonstrated the reduced error in the reproduction of a scene that can be achieved using these filters, compared to an example dye filter set.

I developed the filter sets using only CMOS compatible materials to enable direct integration onto CMOS image sensors in industrial foundries. The filters can be implemented, in-line with the progression of the image sensor manufacturing technology, to replace the current filtering techniques. In the remainder of this Chapter the results presented in this Thesis are summarised and suggestions for future directions are indicated.

7.1 Trichromatic Filters

The colour sensitivity of current imaging systems is difficult to tune because they are limited by the material properties of the dyes used in the filters. For this reason it is difficult to improve the colour reproduction of such systems. The application of plasmonic filters offers greater flexibility in tuning the spectral response, using such filters I was able to design trichromatic filters that mimicked the response of the human eye's daytime colour response. This was a specific example of a more general capability to automate the design of plasmonic filters with arbitrary transmission spectra. I controlled the design process to produce designs that used the same layer thicknesses. This enabled all three filters to be fabricated in a single lithography step.

The simulation of the optimised filter set identified the transmission spectra of the filters with FWHM of 81, 75 and 61 nm, for the red, green and blue filters

respectively. This can be compared to the FWHM of 79, 100 and 55 nm for the corresponding functions. The red filter central wavelength matched exactly that of the corresponding colour matching function. The green and blue had very small red shifts of 8 and 7 nm from the desired central wavelength. The LMS error between the designed filters and the desired spectra were 2.47×10^{-3}, 0.91×10^{-3} and 1.33×10^{-3} for the red green and blue filters respectively. Notably the red filter demonstrated a double peak corresponding to the $\bar{x}(\lambda)$ function spectrum. Such double peak behaviour cannot be achieved using traditional dye filter methods. It can be identified from these results that the optimisation procedure was successful in the design of filters with spectra closely matching that of the colour matching functions.

My colleague Qin Chen fabricated the designs onto a glass microscope slide in a single lithography cycle and I measured the transmission spectra. My experimental filter spectra had FWHM of 117, 150 and 100 nm and transmission coefficient of $25 - 35\%$. The LMS error between the measured filters and the desired colour matching functions were 6.37×10^{-3}, 2.34×10^{-3} and 11.10×10^{-3} for the red, green and blue filters respectively. The increase in the blue filter LMS error is attributed to the slight blue shift in the central wavelength.

I developed a virtual method for quantifying the imaging performance of the filters. Using this I demonstrated the reduced error in the reproduction of a scene imaged using the designed plasmonic filters, compared to that achieved with an example dye filter set. The improvement was shown in terms of both colour error metrics (calculation of the colour error using standard CIE formulae) and human perception (by comparison of colour images reproduced via the various filter sets).

By designing the filter spectra to match the colour matching functions, the error introduced in the conversion of the imager channel signals to the standard CIE system, for the storage and reproduction of colour, was reduced. The application of such a filter set will improve the accuracy of colour conversion and therefore reduce the error in colour reproduction. This optimisation procedure, although specifically applied to the design of plasmonic filters, has the capability to be applied to other electromagnetic design problems (although it requires editing of the design parameters and the simulation input files). An array of the proposed plasmonic filters can be fabricated in a single lithography cycle using

CMOS standard materials. This reduces the cost and fabrication demands from that of dye based filter technology currently applied in CMOS imagers. The filter design presented could in future be incorporated into the lower metal interconnect layers of a standard CMOS imager, or alternatively on top of the silicon of a BSI imager as discussed in Section 2.6.1.1, to reduce the optical crosstalk.

7.2 Multispectral Filters

There has been a demand for cheaper, more portable multispectral imaging systems due to the increased incorporation in applications as diverse as art conservation, crop monitoring and medical imaging. Incorporating an array of multispectral filters on-chip within a CMOS process, similar to current trichromatic imagers, could enable cheaper, mobile, single exposure, multispectral imaging systems. In this work I proposed an alternative design using nanostructuring of an FP cavity to produce multiple filters with different passbands in a single lithography cycle. I presented two designs, one with backfilling of the nanostructured cavity and one without, to alter the effective medium of the cavity via lithographic means.

For the filter design without backfilling I predicted a set of sixteen narrowband filters, with FWHM of $19-77$ nm, could be achieved using only five different cavity lengths. This would require only five lithography cycles which is a significant reduction from the fifteen that would be required using the traditional approach of variation of the physical cavity length alone. The production of filters using this technique was demonstrated experimentally. These results demonstrated that even with a perforated upper mirror, the FP cavity was still operated successfully as a narrowband filter. Using air-filled nanostructuring, as the hole size increased, the reflectivity of the upper mirror of the cavity is reduced degrading filter performance. I presented a theoretical model quantifying the reduction in reflectivity with increasing hole size. This can be applied to aid in the analytical design of such filter sets if the minimum acceptable reflectivity performance required for the application is known.

Using the backfilled design I predicted in simulation that a full set of filters covering the visible range can be achieved using two different physical cavity

lengths. Using my backfilled design Qin Chen fabricated a set of narrowband filters. By using a high-index material a set of twenty three narrowband filters, with a FWHM of $22 - 46\,\text{nm}$, was produced using a single lithography cycle and cavity length. Using a high-index material produces multiple passbands within the visible band. If a set of filters covering the full visible band is desired, an additional highpass filter can be applied. I also applied this filter set within the image reproduction process to demonstrate the reduced colour error that can be achieved with multispectral imaging.

The simulation and experimental filter sets I have presented in this work demonstrate the validity of the nanostructured FP cavity design, and also the application of EMT to represent the cavity, for the production of multiple narrowband filters. I have shown that both designs enable the production of a narrowband filter set, covering the full visible range. By reducing the number of lithography cycles from previous techniques my designs reduce the fabrication complexity. I ensured both filter sets used only materials fully compatible with the CMOS process. Application of the filter designs presented in this work would enable the fabrication of arrays of multispectral filters across an image sensor. Such on-chip multispectral arrays would enable high resolution, simultaneous detection of the full visible spectrum in a compact system. The availability of such an imager would increase the accessibility of multispectral imaging.

7.3 Future Work

The focus of this work has been the development of filters for application within CMOS imagers. The natural progression from the work presented here would be the implementation of the filter designs on-chip. The fabrication of plasmonic filters on-chip has previously been demonstrated [121], this however did not extend to the application of the plasmonic filters in an array for colour imaging. This therefore would be the next suggested progression in the demonstration of this technology. Ultimately the ambition for such filters would be the integration within the chip-stack. The application of the filters from this work, on-chip, would enable the evaluation of the colour reproduction performance for a full colour matching filter based imaging system. Using such an imaging system a di-

rect comparison to the colour reproduction performance of a dye filter technology filter system could be conducted.

As shown in Figure 2.8, the pixel stack of a CMOS imager consists of multiple layers of dielectric materials, this results in oscillations in the transmission spectra due to FP resonances between the layers. In addition, the sensing diode will also have a spectral response that may influence the response of the imaging system. If colour matching function filters are to be produced on-chip, it may be advantageous to include the spectral response of the imaging system used within the advanced design process. This would enable the filter design to be optimised to produce the desired spectra in consideration of the full system.

Similarly, further work could be conducted to explore the application of the FP filters in an array, and their application within an imaging system demonstrated. The filters could be explored to identify if there is any crosstalk directly between filters, or at the pixel floor, that may influence the filtering performance when arrayed. Methodologies, such as optimum arrangements of filters in the array, or barriers between filters, could be developed if such issues exist.

In addition an investigation into the tolerance of the filters to random defects in the nanostructures could be conducted. An exploration of alternative material systems, or further improvement to the design, could also identify if the maximum transmission of both filter techniques could be increased.

The fabrication of the nanostructured FP cavity using nano-imprint or print transfer technology could also be explored. If the filters arrays could be fabricated using such techniques, this would further reduce the fabrication costs and enable incorporation into current imaging systems in BEOL processes.

Appendix A

Implementation of the Drude + 2 Critical Points Model in TEMPEST.

Some materials, such as aluminium as used in this work, have a significant dispersive behaviour in the visible range; the velocity of the propagating wave within the medium will change depending on the wavelength of the excitation. This is a consequence of the wavelength dependent interaction of the waves with the electrons in the medium. This same interaction can also result in attenuation of the wave [122]. The behaviour is often represented by a complex wavelength dependent dielectric function ($\varepsilon(\omega)$).

Although the electromagnetic solver tool TEMPEST had previously been modified to enable broadband illumination in a single simulation, via a Gaussian pulse source, the ability to model a material's dispersive behaviour (in response to the range of wavelengths excited) was not available. The materials were instead implemented with constant n, κ values representing the complex index of refraction at all wavelengths (where $\tilde{n} = n + j\kappa = \sqrt{\varepsilon}$). Since the FDTD simulation is a time-domain method, the time response of the materials must be modelled so the medium behaves appropriately in response to the broadband source, therefore allowing the FDTD simulation to define the structure's behaviour appropriately for the range of incident radiation. For this reason the ability to model the dis-

persive dielectric function of materials was introduced within TEMPEST as part of this work. This involved the identification of a suitable mathematical model to represent the frequency dependent permittivity in an analytical form, then transforming this to a suitable time-domain relation that can be implemented within the standard FDTD algorithm.

A.1 Model Selection

The accuracy of the FDTD simulation of any structure strongly relies on the accuracy of the model implemented to represent the material properties for the wavelength range of interest. For materials such as SiO_2 and Si, models such as the Debye, Drude or Lorentz models can be utilised to define this behaviour [73]. However for metals such as aluminium and gold, the behaviour is more complex, with multiple interband absorptions present in the visible wavelengths, as can be seen in Figure A.1. For such features, a more complicated model is required, with multiple terms [123, 124]. One model presented with improved accuracy across the visible spectra is the Drude + 2 Critical Points (D2CP) model (Equation A.1) [125]. This introduces so called 'critical point' terms to appropriately replicate the complex dispersive behaviour (as shown in Figure A.1). Parameters obtained experimentally [110] are published [124] and are shown in Table reftab:modelpara.

This model can also be applied to model gold, silver and chromium [126]. In [126] it is shown that for silver the D2CP model provides a superior fit to experimental data compared to the previously presented Drude with 4 Lorentz terms [123], with the additional benefit that it requires less computational effort as a result of having fewer terms. This model can be applied to several dispersive metals, with high accuracy, by varying only the input model parameters. The equation is:

$$\varepsilon_{\mathrm{DCP}}(\omega) = \varepsilon_\infty - \frac{\omega_D^2}{\omega^2 + j\gamma\omega} + A_1\Omega_1\left(\frac{e^{j\phi_1}}{\Omega_1 - \omega - j\Gamma_1} + \frac{e^{-j\phi_1}}{\Omega_1 + \omega + j\Gamma_1}\right) + A_2\Omega_2\left(\frac{e^{j\phi_2}}{\Omega_2 - \omega - j\Gamma_2} + \frac{e^{-j\phi_2}}{\Omega_2 + \omega + j\Gamma_2}\right). \tag{A.1}$$

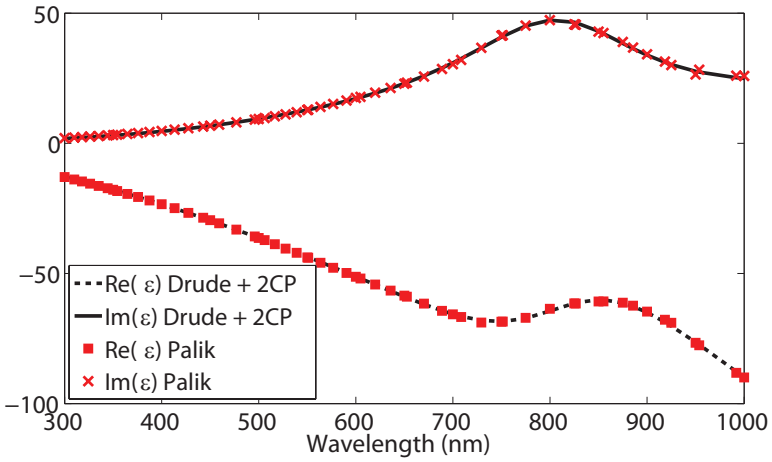


Figure A.1: The Drude + 2 critical points model response using parameters as shown in Table A.1. This is compared to measured permittivity data for aluminium as detailed in [110].

Table A.1: D2CP model parameters for aluminium [124].

ε_∞	1
ω_D	$2.0598e16$
γ	$2.2876e14$
A_1	5.2306
ϕ_1	-0.51202
Ω_1	$2.2694e15$
Γ_1	$3.2867e14$
A_2	5.2704
ϕ_2	0.42503
Ω_2	$2.4668e15$
Γ_2	$1.7731e15$

A.2 Implementation

There are two ways to implement this model into the 3D FDTD algorithm: the recursive convolution method (RC)[127] and the auxiliary differential equation method (ADE)[73]. These methods are generally applicable to a broad variety of materials and models. The RC method was selected because it is the most straightforward to implement, and is more computationally efficient than the

ADE method [128]. The RC process and its derivation is detailed extensively in literature [127, 128]. A detailed description is beyond the scope of this work, however an overview of the theory and implementation is presented here.

The RC approach is based on the description of the electric flux density $\vec{D}(t)$ as a convolution between the electric field intensity and the frequency dependent susceptibility (χ) at any point in a linearly dispersive medium.

$$\vec{D}(t) = \varepsilon_\infty \varepsilon_0 \vec{E}(t) + \varepsilon_0 \int_{\tau=0}^{t} \vec{E}(t-\tau)\chi(\tau)\,\mathrm{d}\tau \tag{A.2}$$

This relation and Equation 3.1 can be combined in their discrete-time form to produce an alternative E field update equation (Equation A.3) to that presented in Section 3.1.1. The RC method approximates the continuous time function $\vec{E}(t)$ as a constant over the time-step Δt and assumes all fields are zero for t less than 0.

$$\begin{aligned}\vec{E}|^{n+1} =& \frac{\varepsilon_\infty}{\varepsilon_\infty + \chi^0}\vec{E}|^n + \frac{\Delta t}{\Delta\varepsilon_0(\varepsilon_\infty + \chi^0)}\vec{\nabla}\times\vec{H}|^{n+1/2} \\ &+ \frac{1}{\Delta\varepsilon_0\left(\varepsilon_\infty + \chi^0\right)} Re\left(\vec{\psi}_D|^n + \sum_{m=1}^{2}\vec{\psi}_m|^n\right)\end{aligned} \tag{A.3}$$

Equation A.3 has several new parameters present, $\chi_0, \varepsilon_\infty$ and the time dependent recursive accumulator, with a component for the Drude term ($\vec{\psi}_D|^n$) and each of the CP ($\vec{\psi}_m|^n$). These are evaluated at the same position as the electric field in the FDTD mesh. For the Drude component this is evaluated as

$$\vec{\psi}_D|^n = e^{-\gamma\Delta t}\vec{\psi}_D|^{n-1} - \left(\frac{\omega_D}{\gamma}\right)^2\left(1 - e^{-\gamma\Delta t}\right)^2\vec{E}|^n, \tag{A.4}$$

and for the CP components

$$\vec{\psi}_m|^n = e^{(-\Gamma_m + j\Omega_m)\Delta t}\vec{\psi}_m|^{n-1} - j\left(\frac{2A_m\Omega_m e^{-j\phi_m}}{\Gamma_m - j\Omega_m}\right)\left(1 - e^{(-\Gamma_m+\Omega_m)\Delta t}\right)^2\vec{E}|^n, \tag{A.5}$$

where n stands for the time instant $t = n\Delta t$. χ^0 in Equation A.3 is given by

$$\begin{aligned}\chi^0 =& \chi_D^0 + Re\left(\sum_{m=1}^{2}\widehat{\chi_m^0}\right)\\ =& -\left(\frac{\omega_D}{\gamma}\right)^2\left(1-e^{-\gamma\Delta t}\right)+\frac{\omega_D^2}{\gamma}\Delta t\\ &+Re\left(\sum_{m=1}^{2}-j(\frac{2A_m\Omega_m e^{-j\phi_m}}{\Gamma_m-j\Omega_m})(1-e^{(-\Gamma_m+\Omega_m)\Delta t})\right).\end{aligned} \tag{A.6}$$

The RC method enables efficient updating of the electric field without needing to perform the initial convolution. These terms are derived from the D2CP model parameters [129]. This implementation requires only the additional storage of the Drude and CP recursive accumulator components, from the previous time step, for each E field component. From this, the summation term is calculated recursively, and can be applied within the new FDTD update equation as outlined in Equation A.3. This also requires the calculation of χ^0 from the model and simulation parameters on initiation. This capability was introduced within the TEMPEST tool by the inclusion of an additional function within the standard TEMPEST algorithm. This is applied by also encoding a newly defined 'dispersive' material type. When a material is defined as dispersive within the input file the new function is called, invoking the alternative update equation for all the cells of that material. On initiation, the tool evaluates the constants and stores these for implementation within the update equation. It should be noted that this new material type accounts for the dispersive electric permittivity only. The materials of interest to this work are non-magnetic materials therefore for the new dispersive material implementation it is assumed $\mu_r = 1$ for all wavelengths. As μ_r is constant the non-dispersive material magnetic field update equation can be applied for the new dispersive material with the incorporation of the new E value as derived from Equation A.3.

The implementation of dispersive materials also impacted other aspects of the TEMPEST algorithm. For example with the introduction of broadband simulations, the calculation of the minimum cell size, according to Equation 3.9, needed to be updated. This implementation of the calculation of the minimum cell size

needed to be altered to account for the fact the wavelength of the source and the material refractive index, was no longer constant for all simulation conditions. Similarly the selection of the wavelength applied in the calculation of the refractive index plot of the structure when applying an 'index' monitor, had to be updated. For this reason if the user does apply the new 'dispersive' material they must also identify the maximum and minimum wavelength (in μm) that is explored in the simulation. The new material is implemented within the input file using an input similar to the standard TEMPEST format [75]. The standard format for a rectangle bounded in the x-direction by [*xl, xh*], in the y-direction by [*yl, yh*], and in the z-direction by [*zl, zh*] is as follows:

> **rectangle** *unit xl xh yl yh zl zh material arg1 arg2 ....*

For example a $10 \times 20 \times 30$ cell rectangle block of SiO_2 ($n = 1.45, k = 0$), would be defined as

> rectangle node 0 10 0 20 0 30 index 1.45 0

The TEMPEST input file code for the new D2CP material is of format:

> **rectangle** *unit xl xh yl yh zl zh dispersive* ε_{inf} ω_D A_1 ϕ_1 Ω_1 Γ_1 A_2 ϕ_2 Ω_2 Γ_2 λ_{min} λ_{max}.

For example a $10 \times 20 \times 30$ cell rectangle block of aluminium, with a broadband pulse excitation between $400 - 750$ nm, would be defined as

> rectangle node 0 10 0 20 0 30 dispersive 1 2.0598e16 2.2876e14 5.2306 -0.51202 2.2694e15 3.2867e14 5.2704 0.42503 2.4668e15 1.7731e15 0.4 0.75

A.3 Validation of Model Implementation

Each update to any simulation code must be validated. The suitability and implementation of this model was evaluated by simulating the transmission intensity through a 20 nm thick aluminium layer in air. The resulting output of the new D2CP aluminium material implemented in a 3D TEMPEST simulation, across

the range of wavelengths of interest, was explored and compared to the analytical solution achieved via the T-matrix method using both experimental data and the D2CP model values of $\varepsilon(\omega)$. The results are plotted in Figure A.2(a) and as can be seen the model and implementation both analytically and in TEMPEST provided less than 4% error to that achieved by applying the experimentally measured values of n, κ (Figure A.2(b)). The error between the analytical and TEMPEST implementation of the D2CP model was less than 2%.

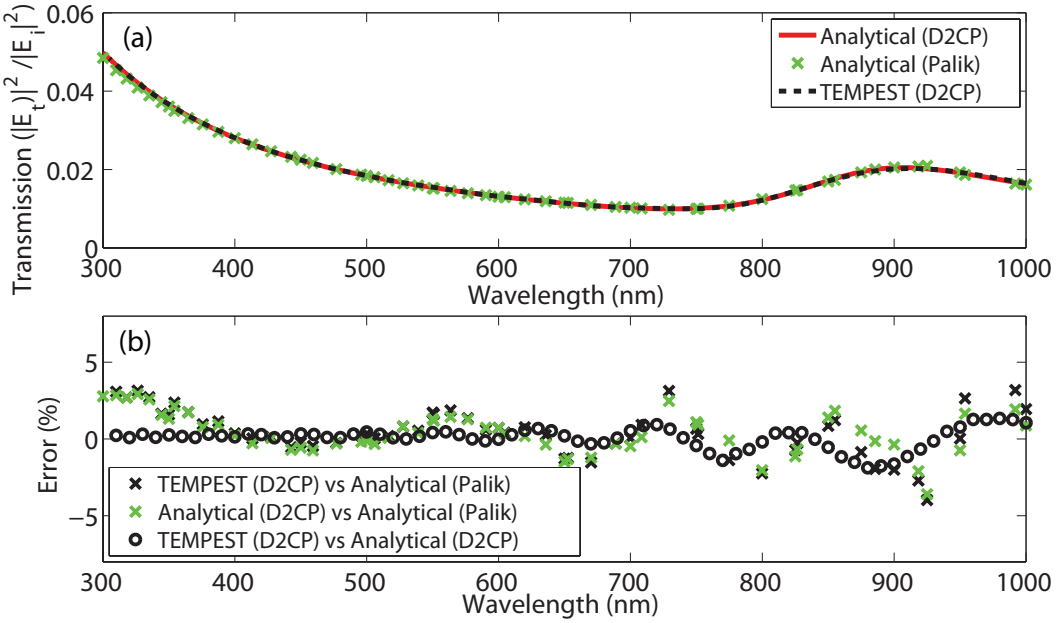


Figure A.2: Implementation of the D2CP model; (a) the transmission through a 20 nm thick aluminium layer, evaluated using TEMPEST with the D2CP model and T-Matrix method for both the D2CP model and the measured ε from [110], and (b) the error in transmission between the different methods.

References

[1] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectronics Journal*, vol. 37, no. 5, pp. 433–451, 2006. 1

[2] CMOS Image Sensors Regain Growth Momentum. IC Insights. Accessed: September 2012. [Online]. Available: http://www.icinsights.com/news/bulletins/CMOS-Image-Sensors-Regaining-Growth-Momentum/ 1

[3] R. Bala, "What to Lose Sleep Over in Digital Color Imaging," *International Congress in Imaging Science: Special Session on Frontiers in Imaging*, May. 2006. 2

[4] R. W. G. Hunt, *The Reproduction of Colour*, 6th ed. Chicester, UK: John Wiley and Sons Ltd, 2004. 2, 8

[5] B. A. Wandell, *Foundations of Vision*. Sunderland, MA: Sinauer Associates, 1995. 2, 7, 8

[6] H. Liang, "Advances in Multispectral and Hyperspectral Imaging for Archaeology and Art Conservation," *Applied Physics A*, vol. 106, no. 2, pp. 309–323, 2012. 2, 38, 83

[7] C. Fischer and I. Kakoulli, "Multispectral and Hyperspectral Imaging Technologies in Conservation: Current Research and Potential Applications," *Studies in Conservation*, vol. 51, no. 2, pp. 3–16, June 2006. 2, 38, 83

[8] K. Martinez, J. Cupitt, D. Saunders, and R. Pillay, "Ten Years of Art Imaging Research," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 28–41, 2002. 2, 38, 83

[9] M. Yamaguchi, M. Mitsui, Y. Murakami, H. Fukuda, N. Ohyama, and Y. Kubota, "Multispectral Color Imaging for Dermatology: Application in Inflammatory and Immunologic Diseases," in *Proceedings of 13th Color Imaging Conference (Society for Imaging Science and Technology/Society for Information Display, 2005)*, 2005, pp. 52–58. 2, 38

[10] J. Gomez-Sanchis, G. Camps-Valls, E. Molto, L. Gomez-Chova, N. Aleixos, and J. Blasco, "Segmentation of Hyperspectral Images for the Detection of Rotten Mandarins," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, A. Campilho and M. Kamel, Eds. Springer Berlin Heidelberg, 2008, vol. 5112, pp. 1071–1080. 2, 122

[11] G. Polder and G. van der Heijden, "Measuring Ripening of Tomatoes Using Imaging Spectrometry," in *Hyperspectral Imaging for Food Quality Analysis and Control*, D.-W. Sun, Ed. Academic Press, 2010, ch. 12, pp. 369–402. 2, 38, 83, 122

[12] A. Picon, O. Ghita, P. Whelan, and P. Iriondo, "Fuzzy Spectral and Spatial Feature Integration for Classification of Nonferrous Materials in Hyperspectral Data," *IEEE Transactions on Industrial Informatics*, vol. 5, no. 4, pp. 483–494, 2009. 2, 39

[13] P. Dillon, A. Brault, J. Horak, E. Garcia, T. Martin, and W. Light, "Fabrication and Performance of Color Filter Arrays for Solid-State Imagers," *IEEE Journal of Solid-State Circuits*, vol. 13, no. 1, pp. 23–27, February 1978. 3, 35

[14] E. Reinhard, E. A. Khan, A. O. Akyz, and G. M. Johnson, *Color Imaging: Fundamentals and Applications*. MA, USA: AK Peters, Ltd., 2008, pp. 678–681. 3, 22, 24, 26, 34, 35

[15] R. W. Sabnis, "Color Filter Technology for Liquid Crystal Displays," *Displays*, vol. 20, pp. 119–129, 1999. 3, 35

[16] R. Ramanath, W. E. Snyder, Y. Yoo, and j. Drew, Mark S, "Color Image Processing Pipeline." 10, 23, 115, 116, 117

[17] S. Westland and C. Ripamonti, *Computational Colour Science Using MATLAB*. Wiley, April 2004. 11, 19

[18] C. McCamy, H. Marcus, and J. Davidson, "A Color Rendition Chart," *J. App. Photographic Engineering*, vol. 2, no. 3, pp. 95–99, 1976. 14

[19] P. Green and L. MacDonald, Eds., *Colour Engineering: Achieving Device Independent Colour*. Chicester, UK: John Wiley and Sons Ltd, 2002, vol. 30, pp. 422–425. 14

[20] A. H. Munselll, *Munsell Book of Color-Matte Finish Collection*. Munsell Color Company, 1976. 14

[21] P. Pellegri, G. Novati, and R. Schettini, "Training Set Selection for Multispectral Imaging System Characterization," *The Journal of Imaging Science and Technology*, vol. 48, no. 3, pp. 203–210, 2004. 14

[22] SharkD. sRGB Approximation of the 1943 Renotations of the Colors Appearing in the 1929 Munsell Book of Color. Accessed: 20th April 2013. [Online]. Available: http://en.wikipedia.org/wiki/File:Munsell_1929_color_solid_transparent.png 14

[23] J. Hiltunen. Database - Munsell Colors Matt (Spec). Accessed: 20th April 2013. [Online]. Available: http://cs.joensuu.fi/~spectral/databases/download/munsell_spec_matt.htm 15

[24] Color Checker RGB and Spectra (20 charts). BabelColor Company. Accessed: 23rd April 2012. [Online]. Available: http://www.babelcolor.com/main_level/ColorChecker.htm#ColorChecker_data 15, 114

[25] G. D. Finlayson and M. S. Drew, "Constrained Least-squares Regression in Color Spaces," *Journal of Electronic Imaging*, vol. 6, no. 24, pp. 484–493, Oct. 16

[26] E. Reinhard, E. A. Khan, A. O. Akyz, and G. M. Johnson, *Color Imaging: Fundamentals and Applications*. MA, USA: AK Peters, Ltd., 2008, pp. 449–461. 17, 21

[27] D. Pascale, "RGB Coordinates of the Macbeth ColorChecker," June 2006. 17, 18

[28] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations. ," *Color Research and Application*, vol. 30, no. 1, pp. 21–30, February 2005. 19, 21, 115

[29] A. E. Gammal and H. Eltoukhy, "CMOS Image Sensors," *IEEE Circuits and Devices Magazine*, May/June 2005. 21, 26

[30] E. R. Fossum, "CMOS Image Sensors: Electronic Camera on a Chip," *IEEE Transactions on Electron Devices*, vol. 44, no. 10, pp. 1689–1698, October 1997. 21, 26

[31] A. Hoffman, M. Loose, and V. Suntharalingham, "CMOS Detector Technology," in *Scientific Detectors for Astronomy*, J. Beletic, J. Beletic, and P. Amico, Eds. Springer Netherlands, 2006, pp. 377–402. 21

[32] E. Hecht, *Optics* , 4th ed. London : Addison-Wesley, 2002. 22, 48, 81, 87

[33] C. Posch, "Detectors, Pixels, and Signal Processing ," in *Smart Cameras*, A. N. Belbachir, Ed. Springer, 2010, ch. 4, pp. 53–79. 22, 26

[34] F. Xiao, J. E. Farrell, and B. A. Wandell, "Psychophysical Thresholds and Digital Camera Sensitivity: The Thousand-Photon Limit," in *Digital Photograph*, 2005, pp. 75–84. 22

[35] J. Vaillant, A. Crocherie, F. Hirigoyen, A.Cadien, and J. Pond, "Uniform Illumination and Rigorous Electromagnetic Simulations Applied to CMOS Image Sensors," *Optics Express*, vol. 15, no. 9, pp. 5494–5503, April 2007. 23, 25

[36] G. Agranov, R. Mauritzson, J. Ladd, A. Dokoutchaev, X. Fan, X. Li, Z. Yin, R. Johnson, V. Lenchenkov, S. Nagaraja, W.Gazeley, J. Bai, H. Lee, A. D'Anna, and G. De-Amicis, "Pixel Continues to Shrink.... Pixel Development for Novel CMOS Image Sensors ," in *Proc. of 2009 International Image Sensor Workshop*, Bergen, Norway, June 2009. 24, 27

[37] R. Fontaine, "Recent Innovations in CMOS Image Sensors," in *Advanced Semiconductor Manufacturing Conference (ASMC), 2011 22nd Annual IEEE/SEMI*, May 2011, pp. 1–5. 24, 26, 27, 28, 30, 31

[38] G. Agranov, R. Mauritzson, S. Barna, J. Jiang, A. Dokoutchaev, X. Fan, and X. Li, "Super Small, Sub 2μm Pixels for Novel CMOS Image Sensors ," *Proc. of 2007 International Image Sensor Workshop*, June 2007. 24, 26

[39] K. Itonaga, K. Mizuta, T. Kataoka, M. Yanagita, S. Yamauchi, H. Ikeda, T. Haruta, S. Matsumoto, M. Harasawa, T. Matsuda, A. Matsumoto, I. Mizuno, T. Kameshima, I. Sugiura, T. Umebayashi, K. Ohno, and T. Hirayama, "0.9 μm Pitch Pixel CMOS Image Sensor Design Methodology," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, Dec. 2009, pp. 1–4. 25, 30, 31

[40] S. Sze, *Semiconductor Devices : Physics and technology*, 2nd ed. John Wiley & Sons, 2002. 26, 32, 33

[41] T. Lulé, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer, and M. Bohm, "Sensitivity of CMOS Based Imagers and Scaling Perspectives," *Electron Devices, IEEE Transactions on*, vol. 47, no. 11, pp. 2110–2122, 2000. 26

[42] A. J. Theuwissen, "CMOS Image Sensors: State-of-the-art," *Solid-State Electronics*, vol. 52, no. 9, pp. 1401–1406, 2008. 26, 30, 31, 32

[43] C. C. Fesenmaier, Y. Huo, and P. B. Catrysse, "Optical Confinement Methods for Continued Scaling of CMOS Image Sensor Pixels," *Optics Express*, vol. 16, no. 25, pp. 20 457–20 470, 2008. 27

[44] T. Joy, S. Pyo, S. Park, C. Choi, C. Palsule, H. Han, C. Feng, S. Lee, J. McKee, P. Altice, C. Hong, C. Boemler, J. Hynecek, M. Louie, J. Lee, D. Kim, H. Haddad, and B. Pain, "Development of a Production-Ready, Back-Illuminated CMOS Image Sensor with Small Pixels," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, Dec. 2007, pp. 1007–1010. 27, 28

[45] J. Wilson and J. Hawkes, *Optoelectronics: An Introduction*, 3rd ed. Prentice Hall, 1998, pp. 19–22. 27, 29

[46] P. B. Catrysse, "Monolithic Intergration of Electronics and Sub-wavelength Metal Optics in Deep Submicron CMOS Technology," in *Material Research Society Symposium Proceedings*, vol. 869, no. 53, 2005. 27, 38

[47] "International Technology Roadmap for Semiconductors: Interconnect," Tech. Rep., Accessed: 6th Oct 2011. [Online]. Available: http://www.itrs.net/Links/2000UpdateFinal/Interconnect2000final.pdf 31

[48] "International Technology Roadmap for Semiconductors: Interconnect," Tech. Rep., 2004, Accessed: 6th Oct 2011. [Online]. Available: http://www.itrs.net/Links/2004Update/2004_08_Interconnect.pdf 31

[49] "International Technology Roadmap for Semiconductors: Interconnect 2010 Update," Tech. Rep., 2010, Accessed: 6th Oct 2011. [Online]. Available: http://www.itrs.net/Links/2010ITRS/2010Update/ToPost/2010_Update_Overview.pdf 31

[50] "International Technology Roadmap for Semiconductors: Interconnect," Tech. Rep., 2011, Accessed: 6th Oct 2011. [Online]. Available: http://www.itrs.net/Links/2011ITRS/Home2011.htm 31

[51] M. Cohen, F. Roy, D. Herault, Y. Cazaux, A. Gandolfi, J. Reynard, C. Cowache, E. Bruno, T. Girault, J. Vaillant, F. Barbier, Y. Sanchez, N. Hotellier, O. LeBorgne, C. Augier, A. Inard, T. Jagueneau, C. Zinck, J. Michailos, and E. Mazaleyrat, "Fully Optimized Cu based process with dedicated cavity etch for 1.75 μm and 1.45 μm pixel pitch CMOS Image Sensors," in *Electron Devices Meeting, 2006. IEDM '06. International*, Dec. 2006, pp. 1–4. 31, 35

[52] R. McGrath, "Image Sensor Technology," in *Single-Photon Imaging*, P. Seitz and A. J. Theuwissen, Eds., 2011, pp. 27–47. 32

[53] "International Technology Roadmap for Semiconductors: Lithography," Tech. Rep., 2011, Accessed: 6th Oct 2011. [Online]. Available: http://www.itrs.net/Links/2011ITRS/Home2011.htm 32, 33

[54] *TWINSCAN NXE:3300B - Technical Specifications*, Accessed: 9th Oct 2011. [Online]. Available: http://www.asml.com/asml/show.do?lang=EN&ctx=46772&dfp_product_id=842 33

[55] B. E. Bayer, "Color Imaging Array," U.S. Patent 3 971 065, July, 1976. 34

[56] H. Honda, Y. Iida, and Y. Egawa, "High Sensitivity Colour CMOS Image Sensor with WRGB Color Filter Array and Color Separation Process Using Edge Detection," in *Proc. of 2007 International Image Sensor Workshop*, Boise, Idaho, U.S.A, June 2007. 34

[57] Y. Inaba, M. Kasano, K. Tanaka, and T. Yamaguchi, "Degradation-Free MOS Image Sensor With Photonic Crystal Colour Filter," *IEEE Electron Device Letters*, vol. 27, no. 6, June 2006. 35, 36

[58] L. Frey, P. Parrein, J. Raby, C. Pellé, D. Hérault, M. Marty, and J. Michailos, "Color Filters Including Infrared Cut-off Integrated on CMOS Image Sensor," *Opt. Express*, vol. 19, no. 14, pp. 13 073–13 080, Jul 2011. 36, 115, 117

[59] D. L. Gilblom, S. K. Yoo, and P. Ventura, "Real-time Colour Imaging with a CMOS Sensor having Stacked Diodes," in *Proc. of SPIE*, vol. 5210, 2004. 36, 37

[60] G. Langfelder, F. Zaraga, and A. Longoni, "Tunable Spectral Responses in a Colour-Sensitive CMOS Pixel for Imaging Applications," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, November 2009. 36, 115, 117

[61] W. Cai, U. K. Chettiar, H.-K. Yuan, V. C. de Silva, A. V. Kildishev, V. P. Drachev, and V. M. Shalaev, "Metamagnetics with rainbow colors," *Opt. Express*, vol. 15, no. 6, pp. 3333–3341, Mar 2007. 37

[62] P. B. Catrysse and B. A. Wandell, "Integrated color pixels in 0.18- μm complementary metal oxide semiconductor technology," *J. Opt. Soc. Am. A*, vol. 20, no. 12, pp. 2293–2306, 2003. 37

[63] A. K. Tilling, G. J. OLeary, J. G. Ferwerda, S. D. Jones, G. J. Fitzgerald, D. Rodriguez, and R. Belford, "Remote Sensing of Nitrogen and Water Stress in Wheat," *Field Crops Research*, vol. 104, no. 13, pp. 77 – 85, 2007. 38, 83

[64] F. Mendoza and J. Aguilera, "Application of Image Analysis for Classification of Ripening Bananas," *Journal of Food Science*, vol. 69, no. 9, pp. E471–E477, 2004. 38, 83, 122

[65] F. G. Smith and T. A. King, *Optics and Photonics: An Introduction*. Wiley, 2000. 38

[66] G. Bianco, F. Bruno, and M. Muzzupappa, "Multispectral Data Cube Acquisition of Aligned Images for Document Analysis by Means of a Filter-Wheel Camera Provided with Focus Control," *Journal of Cultural Heritage*, 2012. 39

[67] VariSpec Liquid Crystal Tunable Filter Datasheet. Accessed: 21st April 2013. [Online]. Available: http://www.perkinelmer.co.uk/CMSResources/Images/44-140156DTS_010053A_01_VariSpec_DTS.pdf 39

[68] S. Zhi-Xue, L. Jian-Feng, Z. Da-Yong, L. Yong-Quan, L. Yan, H. Li-Xian, L. Hai-Tao, and L. Fei, "Research on LC-Based Spectral Imaging System for Visible Band," *Proceedings of SPIE*, vol. 8181, pp. 1–7, 2011. 40

[69] A. Mitra, H. Harutyunyan, S. Palomba, and L. Novotnyo, "Tuning the Cavity Modes of a Fabry-Perot Resonator Using Gold Nanoparticles," *Optics Letters*, vol. 35, no. 7, pp. 953 – 955, April 2010. 40

[70] A. Mehta, R. C. Rumpf, Z. Roth, and E. G. Johnson, "Nanofabrication of a Space-Variant Optical Transmission Filter," *Optics Letters*, vol. 31, no. 19, pp. 2903–2905, October 2006. 40

[71] G. Shambat, M. S. Mirotznik, G. Euliss, V. O. Smolski, E. G. Johnson, and R. A. Athale, "Photonic Crystal Filters for Multi-band Optical Filtering on a Monolithic Substrate," *Journal of Nanophotonics*, vol. 30, pp. 1–11, March 2009. 41

[72] K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, 1966. 42

[73] A. Taflove and S. C. Hagness, *Computational Electrodynamics: the finite-difference time domain method*, 2nd ed. Artech House Publishers, 2000. 43, 44, 45, 132, 133

[74] *FDTD Solutions Reference Guide: Release 7.5.* [Online]. Available: http://docs.lumerical.com/en/previous_fdtd/FDTD_reference_guide.pdf 46

[75] A. Wong, *TEMPEST Users Guide*, August 1994. 47, 136

[76] A. K.-K. Wong, "Rigorous Three-dimensional Time-Domain Finite-Difference Electromagnetic Simulation," Ph.D. dissertation, September 1994. 47

[77] T. V. Pistor, "Electromagnetic Simulation and Modeling with Applications in Lithography," Ph.D. dissertation, Spring 2001. 47

[78] T. D. Drysdale, "Passive Devices for Terahertz Frequencies," Ph.D. dissertation, University of Canterbury, January 1993. 47, 49, 54

[79] J. H. Davies, *The Physics of Low-Dimensional Semiconductors: An Introduction*, 1st ed. Cambridge University Press, 2008, pp. 153–155. 48

[80] J. M. Johnson and Y. Rahmat-Samii, "Genetic Algorithms in Engineering Electromagnetics," *IEEE Antennas and Propagation Magazine*, vol. 39, no. 4. 51, 52, 53, 54

[81] D. S. Weile and E.Michielssen, "Genetic Algorithm Optimization Applied to Electromagnetics: A Review," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, March 1997. 51, 52, 53, 54

[82] M. Mitchell, *An Introduction to Genetic Algorithms.* MIT Press Massachusetts, 1998. 51, 52

[83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, 1983. 51

[84] W. L. Goffe, G. D. Ferrier, and J. Rogers, "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, vol. 60, 1994. 51

[85] R. S. Sexton, R. E. Dorsey, and J. D. Johnson, "Optimization of Neural Networks : A Comparative analysis of the Genetic Algorithm and Simulated Annealing," *European Journal of Operational Research*, vol. 114, 1999. 51, 54

[86] Z. Li, P. Papalambros, and J. Volakis, "Frequency Selective Surface Design by Integrating Optimisation Algorithms with Fast Full Wave Numerical Methods," *IEEE Proc. Antennas and Propagation*, vol. 149, no. 3, June 2002. 51

[87] M. Ohira, H.Deguchi, M.Tsuji, and H.Shigesawa, "Multiband Single-Layer Frequency Selective Surfaces Designed by Combination of Genetic Algorithm and Geometry-Refinement Technique," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 11, November 2004. 51

[88] G. Manara, A.Monorchio, and R.Mittra, "Frequency Selective Surface Design Based on Genetic Algorithm," *Electronics Letters*, vol. 35, no. 17, August 1999. 51

[89] B. Sareni and L. Krahenbuhl, "Fitness Sharing and Niching Methods Revisited," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 3, September 1998. 54

[90] L. Gao and Y. Hu, "Multi-target Matching based on Niching Genetic Algorithm," *International Journal of Computer Science and Network Security*, vol. 6, no. 7A, July 2006. 54

[91] S. Cui and D. S. Weile, "Robust Design of Absorbers Using Genetic Algorithms and the Finite Element-Boundary Integral Method," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 12, December 2003. 54

[92] J. Tominaga, "Localized Surface Plasmons for Optical Data Storage Beyond the Diffraction Limit," in *Surface Plasmon Nanophotonics*, ser. Springer Series in Optical Sciences, M. Brongersma and P. Kik, Eds. Springer Netherlands, 2007, vol. 131. 59

[93] Q. Chen and D. R. S. Cumming, "Visible Light Focusing Demonstrated by Plasmonic Lenses Based on Nano-slits in an Aluminum Film," *Opt. Express*, vol. 18, no. 14, pp. 14 788–14 793, Jul 2010. 59

[94] X. Zhang, M. Ambati, N. Fang, H. Lee, Z. Liu, C. Sun, and Y. Xiong, "Optical superlens," in *Surface Plasmon Nanophotonics*, ser. Springer Series in Optical Sciences, M. Brongersma and P. Kik, Eds. Springer Netherlands, 2007, vol. 1312. 59

[95] S. A. Maier, *Plasmonics: Fundamentals and Applications*. Springer, 2007. 59, 60, 61, 62, 63, 64

[96] V. M. S. Vladimir P. Drachev, Mark D. Thoreson, "Sensing Proteins with Adaptive Metal Nanostructures," in *Surface Plasmon Nanophotonics*, ser. Springer Series in Optical Sciences, M. Brongersma and P. Kik, Eds. Springer Netherlands, 2007, vol. 131. 59

[97] T. A. Klar, "Biosensing with Plasmonic Nanoparticles," in *Nanophotonics with Surface Plasmons*, V. M. Shalaev and S. Kawata, Eds. Elsevier Science, 2007, ch. 8, pp. 219–270. 59

[98] A. D. William L. Barnes and T. W. Ebbesen, "Surface Plasmon Subwavelength Optics," *Nature*, vol. 424, pp. 824–830, August 2003. 60, 62

[99] T. W. Ebbesen, H. Lezec, H. Ghaemi, T. Thio, and P. Wolff, "Extraordinary Optical Transmission Through Sub-wavelength Hole Arrays," *Nature*, vol. 391, no. 6668, pp. 667–669, 1998. 63

[100] A. Krishnan, T. Thio, T. Kim, H. Lezec, T. Ebbesen, P. Wolff, J. Pendry, L. Martin-Moreno, and F. Garcia-Vidal, "Evanescently Coupled Resonance in Surface Plasmon Enhanced Transmission," *Optics Communications*, vol. 200, pp. 1–7, 2001. 63

[101] V. Rivera, F. Ferri, O. Silva, F. Sobreira, and E. Marega Jr, "Light Transmission via Subwavelength Apertures in Metallic Thin Films," in *Plasmonics - Principles and Applications*, K. Y. Kim, Ed. Intech, 2012, ch. 7, pp. 157–1822. 64

[102] W. Zhou, H. Gao, and T. W. Odom, "Toward Broadband Plasmonics: Tuning Dispersion in Rhombic Plasmonic Crystals," *ACS Nano*, vol. 4, no. 2, pp. 1241–1247, 2010. 64

[103] H. Ghaemi, T. Thio, D. e. a. Grupp, T. W. Ebbesen, and H. Lezec, "Surface Plasmons Enhance Optical Transmission Through Subwavelength Holes," *Physical Review B*, vol. 58, no. 11, 1998. 64

[104] C.Genet and T.W.Ebbensen, "Light in Tiny Holes," *Nature*, vol. 445, January. 65

[105] F. Przybilla, A. Degiron, J.-Y. Laluet, C. Genet, and T. W. Ebbesen, "Optical Transmission in Perforated Noble and Transition Metal Films," *Journal of Optics A: Pure and Applied Optics*, vol. 8, no. 5, p. 458, 2006. 65

[106] D. Pacifici, H. J. Lezec, L. A. Sweatlock, R. J. Walters, and H. A. Atwater, "Universal Optical Transmission Features in Periodic and Quasiperiodic Hole Arrays," *Opt. Express*, vol. 16, no. 12, pp. 9222–9238, Jun 2008. 65

[107] A. Degiron, H. Lezec, W. Barnes, and T. Ebbesen, "Effects of Hole Depth on Enhanced Light Transmission Through Subwavelength Hole Arrays," *Applied Physics Letters*, vol. 81, no. 23, pp. 4327–4329, 2002. 65

[108] Q. Chen and D. R. S. Cumming, "High Transmission and Low Color Crosstalk Plasmonic Color Filters Using Triangular-lattice Hole Arrays in Aluminium Films." *Optics Express*, vol. 18, no. 13, pp. 14 056–14 062, June 2010. 65, 66, 67, 79

[109] Q. Chen, C. Martin, and D. R. S. Cumming, "Transfer Printing of Nanoplasmonic Devices onto Flexible Polymer Substrates from a Rigid Stamp," *Plasmonics*, pp. 1–7. 75

[110] E. D. Palik, Ed., *Handbook of Optical Constants of Solids*, 2nd ed. Academic, 1985. 80, 132, 133, 137

[111] M. Hassan, "Theory of the Fabry-Perot Interferometer with Nonparallel Reflectors," *Soviet Physics Journal*, vol. 10, no. 9, pp. 41–44, 1967. 84

[112] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. Cambridge University Press, 1999. 85, 87

[113] G. Hernandez, *Fabry-Perot Interferometers*. Cambridge University Press, 1986, pp. 23–31. 88

[114] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. Cambridge University Press, 1999, pp. 837–839. 88

[115] S. M. Rytov, "Electromagnetic Properties of a Finely Stratified Medium," *Soviet Physics, JETP*, vol. 2, no. 3, pp. 466–475, May 1956. 88, 89

[116] D. R. S. Cumming and R. J. Blaikie, "A Variable Polarisation Compensator Using Artificial Dielectrics," *Optics Communications*, vol. 163, no. 46, pp. 164 – 168, 1999. 90

[117] H. Zhou, C. Sim, A. Glidle, C. Hodson, C. R. Kinsey, and C. D. W. Wilkinson, "Properties of Silicon Nitride by Room-Temperature Inductively Coupled Plasma Deposition," in *Plasma Processes and Polymers*. Wiley-VCH Verlag GmbH & Co. KGaA, 2005, ch. 6, pp. 77–86. 107

[118] Selected Colorimetric Tables. International Commission on Illumination (CIE). Accessed: 11th April 2013. [Online]. Available: http://www.cie.co.at/index.php/LEFTMENUE/index.php?i_ca_id=298 114

[119] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and

Spectrum," Department of Computer Science, Columbia University, Tech. Rep., Nov 2008. [Online]. Available: http://www.cs.columbia.edu/CAVE/databases/multispectral/ 117

[120] S. H. C. G. Baldridge, A. M. and G. Rivera, "The ASTER Spectral Library Version 2.0." *Remote Sensing of Environment*, vol. 113, pp. 711–715, 2009. 122

[121] Q. Chen, D. Das, D. Chitnis, K. Walls, T. D. Drysdale, S. Collins, and D. R. S. Cumming, "A CMOS Image Sensor Integrated with Plasmonic Colour Filters," *Plasmonics*, pp. 1–5, 2012. 129

[122] E. Hecht, *Optics* , 4th ed. London : Addison-Wesley, 2002, pp. 127–130. 131

[123] F. Hao and P. Norlander, "Efficient Dielectric Function for FDTD Simulation of the Optical Properties of Silver and Gold Nanoparticles," *Chemical Physics Letters*, vol. 446, 2007. 132

[124] A. Vial and T. Laroche, "Description of the Dispersion Properties of Metals by Means of the Critical Points Model and the Application to the Study of Resonant Structures Using the FDTD Method," *J. Phys. D: Appl. Phys*, vol. 40, 2007. 132, 133

[125] P. G. Etchegoin, E. C. L. Rue, and M. Meyer, "An Analytical Model for the Optical Properties of Gold," *J. Chem. Phys*, vol. 125, 2006. 132

[126] A. Vial and T. Laroche, "Comparisons of Gold and Silver Dispersion Laws Suitable for FDTD Simulations," *Appl. Phys. B*, vol. 93, 2008. 132

[127] R. J. Luebbers, K. S. Kunz, and F. Hunsberger, "A Frequency-Dependent Finite-Difference Time-Domain Formulation for Transient Propagation in Plasma," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 1. 133, 134

[128] R. J. Luebbers and F. Hunsberger, "FDTD for Nth-Order Dispersive Media," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 11, November 1992. 134

[129] A. Vial, "Implementation if the Critical Points Model in the Recursive Convolution Method for Modelling Dispersive Media with the Finite-Difference Time Domain Method," *Journal of Optics A: Pure and Applied Optics*, vol. 9, 2007. 135